

Article

Development and Analysis of a CNN- and Transfer-Learning-Based Classification Model for Automated Dairy Cow Feeding Behavior Recognition from Accelerometer Data

Victor Bloch ^{1,*}, Lilli Frondelius ¹, Claudia Arcidiacono ^{2,*} , Massimo Mancino ² and Matti Pastell ¹ ¹ Natural Resources Institute Luke (Finland), Latokartanonkaari 9, 00790 Helsinki, Finland² Department of Agriculture, Food and Environment (Di3A), Building and Land Engineering Section, University of Catania, Via Santa Sofia 100, 95123 Catania, Italy

* Correspondence: victor.bloch@luke.fi (V.B.); claudia.arcidiacono@unict.it (C.A.)

Abstract: Due to technological developments, wearable sensors for monitoring the behavior of farm animals have become cheaper, have a longer lifespan and are more accessible for small farms and researchers. In addition, advancements in deep machine learning methods provide new opportunities for behavior recognition. However, the combination of the new electronics and algorithms are rarely used in PLF, and their possibilities and limitations are not well-studied. In this study, a CNN-based model for the feeding behavior classification of dairy cows was trained, and the training process was analyzed considering a training dataset and the use of transfer learning. Commercial acceleration measuring tags, which were connected by BLE, were fitted to cow collars in a research barn. Based on a dataset including 33.7 cow × days (21 cows recorded during 1–3 days) of labeled data and an additional free-access dataset with similar acceleration data, a classifier with F1 = 93.9% was developed. The optimal classification window size was 90 s. In addition, the influence of the training dataset size on the classifier accuracy was analyzed for different neural networks using the transfer learning technique. While the size of the training dataset was being increased, the rate of the accuracy improvement decreased. Beginning from a specific point, the use of additional training data can be impractical. A relatively high accuracy was achieved with few training data when the classifier was trained using randomly initialized model weights, and a higher accuracy was achieved when transfer learning was used. These findings can be used for the estimation of the necessary dataset size for training neural network classifiers intended for other environments and conditions.

Keywords: cow behavior; CNN classifier; acceleration tags; transfer learning; dataset variability; open-source dataset



Citation: Bloch, V.; Frondelius, L.; Arcidiacono, C.; Mancino, M.; Pastell, M. Development and Analysis of a CNN- and Transfer-Learning-Based Classification Model for Automated Dairy Cow Feeding Behavior Recognition from Accelerometer Data. *Sensors* **2023**, *23*, 2611. <https://doi.org/10.3390/s23052611>

Academic Editors: Štefan Kohek and Krista Rizman Zalik

Received: 3 January 2023

Revised: 21 February 2023

Accepted: 22 February 2023

Published: 27 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Farm animal activity recognition is important for livestock health and welfare monitoring. Sensors for the behavioral recognition of dairy cows have been developed and produced for at least two decades [1]. Based on acceleration tags, numerous commercial systems [2–4] can provide high accuracy in behavior recognition. Nevertheless, the commercial systems usually do not provide access to the raw acceleration data, which is highly important for researchers studying the animal behavior and developing new methods for efficient farm management. In addition, the price of the equipment and its maintenance is impractical for small farms or farms with small ruminants.

New sensors are constantly being developed in this research. This has been inspired by new technologies that provide a smaller device size [5], better data transfer possibilities, and a lower energy consumption, such as Bluetooth low energy—BLE [6]. Due to progress made in data processing methods such as deep neural networks, the accuracy and robustness of

the algorithms monitoring the animal behavior have been constantly improved [7]. The machine learning model development process includes data pre-processing (e.g., handling the records with missing data, filtering the raw time series, calculating additional time series, and segmenting the time series into time windows), calculating features for some classifiers, and model training and postprocessing. Riaboff et al. [7] provided an extensive review of these aspects in livestock applications.

In recent years, methods based on convolutional neural networks (CNN) have been widely used for recognition applications such as human activity recognition (HAR) [8,9]. In livestock applications, CNN for acceleration data, which is measured by tags fitted to the animals, has been used by [10–12]. The design of a neural network (NN)-based behavior classifier includes a number of factors. In particular, the type and the architecture of the NN must be fitted to its application. For livestock, CNNs and recurrent NNs (RNNs) with 2–4 convolutional layers have been utilized to process time series data. However, deep learning models, particularly CNNs, are still not used widely [7].

To train deep neural networks effectively, a large amount of reference data must usually be collected and labeled [13]. Different sensors are used to achieve labeled references for the cow behavior and body position: e.g., feeders for estimating the feeding time [14], or halter sensors to measure rumination and feeding time [10]. Manual labeling can be performed from direct cow observation [15] or video recorded by cameras installed in the cow environment [16]; however, this method is highly time-consuming. In cases where observations of actual behavior cannot be made, unsupervised methods for the behavior classification are used [17]. To advance the development of HAR, some researchers published data used for their studies in open access, saving time for reference preparation and enabling the use of larger datasets for model training [18] (wireless sensors data mining—WISDM). According to [7], a large variety of collected data for the classifier training have been used in different studies (from 2 to 200 h), and a recommendation to collect data from at least from 25 animals for at least 40 h was given. An analysis of required training data was performed for some data series [19]. However, no analysis of a required amount of the training data was found for farm animal activity recognition and HAR.

Different types of data augmentation have been used to enlarge the training dataset [12,20]: rotation, permutation, jittering and scaling performed for the original signal, or local averaging as a down-sampling technique and shuffling in the feature space [21]. However, the specific augmentation, as well as the optimization of the classification window length (epoch, time window, segment, observation) was performed in each study for its specific datasets.

Transfer learning is a method that prepares a classification model for one dataset and uses this pretrained model as a base for training a model for another similar dataset. For example, training a pretrained model based on younger population groups and using it as the initial condition to train a model for older people, as was carried out by [19,20]. The additional training of pretrained models using data from specific objects and environments improves their accuracy and decreases the training time relative to newly trained models or existing models. In HAR, this method was used by [22,23]. According to our review, this method has not been used for livestock activity recognition.

In this study, we evaluated the minimal amount of data needed to effectively train an NN classifier and the use of transfer learning based on an openly available dataset. We used a low-cost, open-source system based on acceleration tags to develop a behavior classification method according to the best practices taken from the reviewed studies. The aims of this study were to (a) develop a behavior classification model, (b) evaluate the impact of the size of training dataset and (c) evaluate the use of transfer learning on the accuracy of farm animal activity recognition classification using CNNs.

2. Materials and Methods

2.1. Barn Study Area and Monitored Cows

The data used for the development and validation of the system were collected from dairy cows housed in a free-stall research barn (Luke, Maaninka, Finland) from 4 March till

15 April 2021. The barn comprised two separated sections with a 10×20 m area containing, in total, 48 lying stalls and 24 feeders (Hokofarm, Marknesse, The Netherlands) as shown in Figure 1. A group of 48 cows, specifically Ayrshire ($n = 18$) and Holstein ($n = 30$) cows, were housed in the study area during the lactation period (the average parity was 2.3 with a minimum of 1 and a maximum of 7, the average \pm STD days of lactation were 126 ± 82). The barn was equipped with continuously recording cameras (HDBW5541R, 5M, 3.6 mm, Dahua, Hangzhou, China). The cameras were installed on the ceiling at a height of 6 m and covered a major part of the area. Each section of the barn included a 140 m^2 free area, winter natural ventilation, automatic manure scraping (Lely Discovery 90SW, Lely, Maassluis, The Netherlands), fresh feed delivered six times daily and freely available water.



Figure 1. Top-view image of the research barn acquired by one of the cameras.

2.2. System Design

Tags measuring 3D acceleration (RuuviTag, Ruuvi Innovations, Riihimäki, Finland) were packed in plastic boxes and adjusted by a Velcro belt on the left side of the cow collars or to one of the legs just above the metatarsal joint (Figure 2). In total, 96 tags were fitted to both the collars and legs of 48 cows. The tags attached to the legs were used only to test the reliability of the wireless transmission. The number of transmitting tags was typical for a commercial barn. The tags broadcasted the data as BLE advertising packets. The acceleration was sampled at 25 Hz, and the frequency of the message sending was 5 Hz. Each packet included five samples for three axes, which amounted to 15 acceleration values. The data from the tags were received by six receiving stations, which were single-board computers equipped with a Bluetooth antenna (Raspberry Pi 3 B+, Raspberry Pi Foundation, Cambridge, UK). The stations were packed in hermetic cases with heat sink ribs and installed at 3–5 m height on the barn structures (Figure 2c). They were evenly distributed in the study area to minimize the distance from the tags sufficient for receiving the broadcasted signal [24]. The receiving stations recorded the tag accelerations and the receiving time. The data were stored on the base station's storage and were sent via a local network, which was maintained by a router (EA7500, Linksys), through a message-queuing protocol (ZeroMQ, iMatix Corporation, Brussels, Belgium). A PC (Intel(R) Core(TM) i7-9750H, CPU 2.6 GHz, RAM 16 GB) received the messages and stored the raw data in CSV files. The tag and base station software were written in C++ (version 2020), and the C# language was used (Microsoft, Redmond, WA, USA)

for the PC. The lifetime of the tag battery was estimated using power profiling, which was described in [24] as about three years.

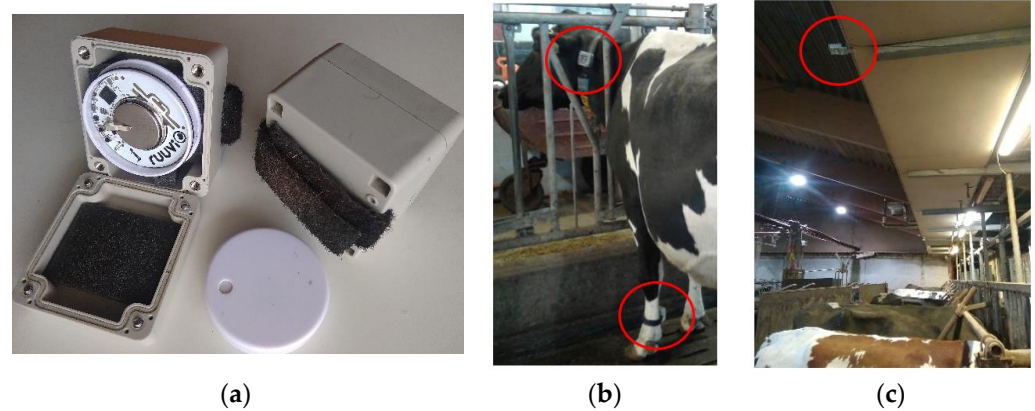


Figure 2. Component of the location and acceleration measuring system installed in a barn: RuuviTag inside a protecting plastic box (a), tag on the cow collar (b) and receiving station installed on a barn structure (c) marked by red circles.

2.3. Data Collection and Labeling

Three feeding behavior classes were considered in this study: feeding, ruminating and other (neither ruminating nor feeding).

The individual feeders were used to collect reference data for the feeding. We assumed that animals were not ruminating while registered to the feeders. Manual labeling was used to recognize ruminating and other behaviors. Individual cows were recognized by their unique coat color patterns; images of cows from both sides, on top and from the head were captured at the beginning of the experiment to aid in recognition. Only time intervals during which the cow and its behavior were clearly detected were labeled. The time labels of the feeders and the cameras were synchronized with the time label of the tag receiving stations with an accuracy of ± 1 s. The labeling was performed by one trained person. An ethogram of behavior classification for visual observations from [4] was used in this study.

The average rate of the missing data messages was $52.6 \pm 6.1\%$ (mean \pm STD). The missing samples were concentrated in groups containing multiplication of five samples (since the data were transferred in packets including five samples), as is shown in Figure 3.

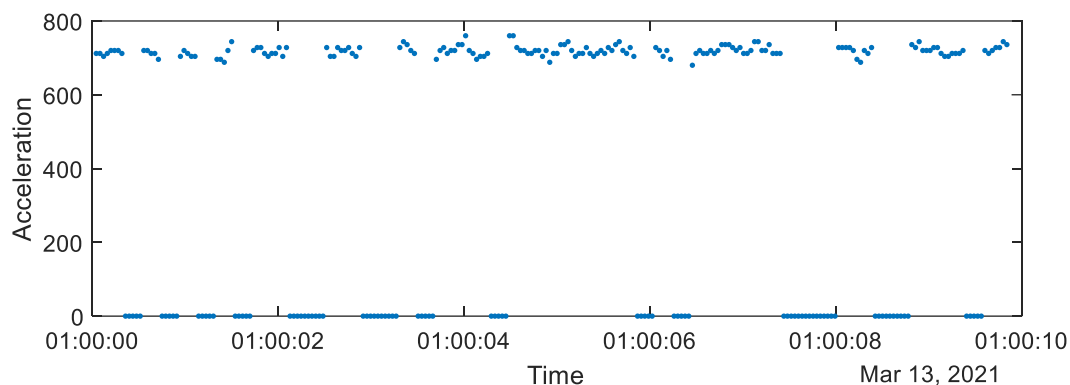


Figure 3. Illustration of the missing samples (value 0) in the recorded acceleration data (dots).

The classification models were trained on two datasets (Table 1): data collected in this study and labeled for 21 cows, as explained above (4–13 March 2021), and open-source data published by [10]. Different from the current study, [10] used sampling at 10 Hz and sensors with the ability to download the data, thus preventing data loss. All data labeling

was performed automatically with the help of halters that measured the cows' feeding behavior (Rumiwatch, Ettenhausen, Switzerland).

Table 1. Characteristics of datasets used for NN model training (mean \pm STD).

	<i>N</i>	Period, Days	Average Time, Hours	Total Time, Hours (Days)	Fe, %	Ru, %	Oth, %
Collected data	21	1–3	38.5 \pm 12.4	809 (33.7)	19.7 \pm 5.7	36.9 \pm 6.1	43.3 \pm 6.9
Open-source data	18	6–18	191.7 \pm 87.5	3450.5 (143.7)	17.6 \pm 3.8	38.4 \pm 3.5	43.9 \pm 6.6

N—number of cows, Fe—feeding, Ru—ruminating, Oth—other behaviors.

2.4. Data Processing

Data pre-processing included filtering, amplitude normalization, sampling frequency normalization, augmentation and balancing. The raw acceleration data from the neck tags was filtered by a Hamming high-pass filter with a filter order of 511 and a cut-off frequency of 0.1 Hz. The acceleration values were normalized to ± 1 . Due to a high rate of missing samples, the missing data were replaced by zeros to preserve the structure of the time series. The data were used for the training according to the methods proposed by [25].

A window overlap augmentation of consecutive windows with a 50% overlap was used to increase the amount of the training data in accordance with the recommendations of [7] and the review by [8]. Since tags fitted on collars were able to rotate around the neck, a rotational augmentation was used to simulate a possible rotation around the X axis parallel to the cows' neck and to train the model to be insensitive to the tag orientation. The Y and Z acceleration components were rotated in a 3D space by the transformation

$$\begin{pmatrix} a_X \\ a_Y \\ a_Z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} a_X \\ a_Y \\ a_Z \end{pmatrix}, \quad (1)$$

where a_X , a_Y and a_Z are the components of the tag acceleration measured along the X, Y and Z axes and α is a random rotational angle, as per several authors [12,18,26]. Every classification window was rotated by a random angle.

To balance data for all behavioral classes, data windows from the minor classes were randomly taken, rotational augmentation on a random angle was performed and the result was added to the training set. The balancing was performed for data collected from each individual cow for one day. An increase in the dataset size due to the balancing depended on the level of imbalance and was, on average, $41 \pm 20\%$.

Postprocessing of the classified behavior was achieved by a median filter with a window length equal to five. Both the data collected for this study and the data published by [10] were processed using the same procedure. Additionally, the sampling frequency of the Pavlovic [10] dataset was changed, using zero padding, from the original 10 Hz to the 25 Hz used in the current system.

2.5. Tested Classifying Models

Two NN classifiers found in the reviewed literature were compared in this study:

CNN2. Methods for the human activity recognition, described in [27]. The CNN2 consists of two 1D convolutional layers with a kernel size of 3, a dropout layer and a pooling layer, as presented in Figure 4.

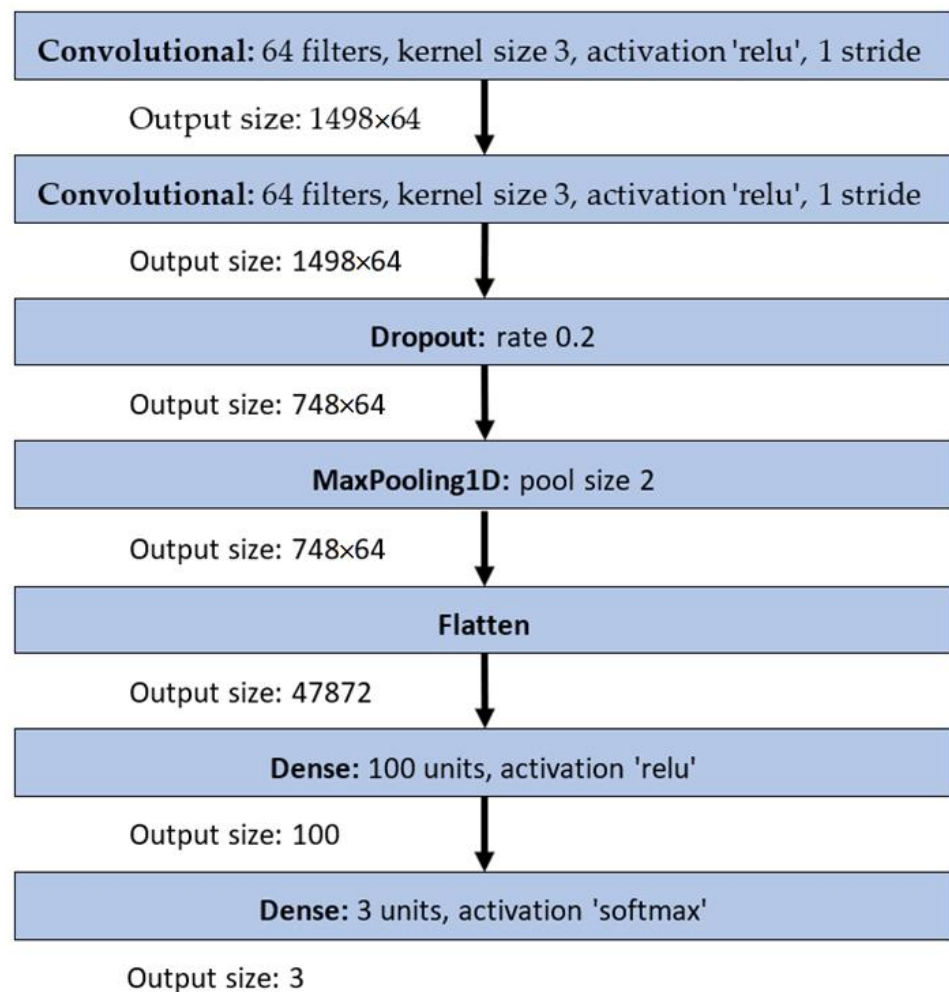


Figure 4. CNN2 architecture.

CNN4. Deep CNN for cow activity recognition, as described in [10]. The CNN4 consists of four 1D convolutional layers with kernel sizes of 52 and 1, a dropout layer and a pooling layer.

The models' structures are available in the Supplementary Materials https://github.com/cowbhav/CowBhave_BehaviorModel (accessed on 2 January 2023).

The size of the classification window was optimized similarly to what found in the reviewed studies [28,29]. In this study, the optimal classification window size was searched by a grid search algorithm in the set [5 10 30 60 90 120 180 300] s with the extreme values 5 s and 300 s, which were taken from Table A1 for cows. The total amount of the bouts for each behavior class in the available labeled dataset with a length less than 300 s was 3.1% for feeding, 2.7% for rumination and 9% for the other behaviors.

The pretrained classification models for the transfer learning for different window sizes were trained by the data published by [10]. The transfer learning was achieved by training the last convolutional layer in the models (the second for the CNN2 and the fourth for the CNN4 model) and all subsequent layers. The training was performed for 30 epochs with a 0.001 learning rate. An Adam optimizer was used. The model was implemented in Python and trained using the Keras library with Tensorflow.

The datasets used for training (which were the dataset collected in this study and the dataset published by [10]) were acquired from different cows and environments with different sensors, sampling rates and rates of missing data. Hence, they were used to test the applicability of the trained models for other cows and environments. The classification accuracy of models trained on one dataset and validated by another dataset was estimated.

2.6. Analysis on the Effect of Training Dataset Size

To evaluate the dependence of the model accuracy on the amount of data used for the model training, learning using randomly initialized model weights and transfer learning were performed using different parts of the original dataset. The smallest data amount which can be used for training is the classification window. For this test, only a 60 s window, including $60 \times 25 = 1500$ acceleration samples (for which 25 Hz was the accelerometer sampling frequency), was used. Each window had its behavior class label; hence, it was defined as a training sample and the dataset size was measured by the number of training samples. Small testing datasets were created as parts of the original dataset. The original acceleration data for one cow for one day was stored in one file, resulting in a total of 56 files. To create the minimal dataset, one training sample for all three classes was taken from each file, totaling $3 \times 56 = 168$ training samples ($3 \times 56 \times 60 \text{ s} = 10,080 \text{ s} = 2.8 \text{ h}$). This amount represents 0.34% of the total dataset size of 48,540 training samples. For the next dataset, including two training samples taken from each file, the dataset size was $3 \times 56 \times 2 = 336$ training samples. Hence, in the analysis, the initial datasets, including the following dataset sizes, were used: 168, 336 . . . 1680, 4854, 9708 . . . 48,540. In larger datasets for which an equal amount of training samples for each behavior class could not be taken, actual data size for the training was increased by resampling. The sizes of the corresponding datasets enlarged by the augmentation and balancing were: 336, 504 . . . 3192, 13,878, 28,358 . . . 132,762. For each fold in the 10-fold validation, the size of dataset was multiplied by a factor equal to approximately 9/10.

2.7. Accuracy Evaluation

The accuracy was estimated by the total classification precision and macro F1, confusion matrix, and (micro) $Precision_i$, $Recall_i$ and $F1_i$ for each separate behavior class, i , as follows in Equation (2):

$$\begin{aligned} Precision_i &= \frac{TP_i}{TP_i + FP_i}, \quad Recall_i = \frac{TP_i}{TP_i + FN_i}, \quad F1_i = 2 \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i}, \quad i = 1, 2, 3; \\ Precision &= \frac{TP}{TP + FP}, \quad F1 = \text{mean}(F1_i)_{i=1,2,3}, \end{aligned} \quad (2)$$

where TP_i is the number of true positive classifications for the class i , FP_i is the number of false positive classifications for the class i , and FN_i is the number of false negative classifications for the class i . TP is the total number of correct classifications for all classes and FP is the total number of incorrect classifications for all classes.

3. Results

The performance of the CNN2 model for the 60 s window size, trained using random weights and by transfer learning for different dataset sizes used for the training, is presented in Figure 5. The use of transfer learning clearly improved the performance (F1 score) of the classifier for small dataset sizes, and the average F1-score 87% was obtained with just 336 training samples. The advantage of using transfer learning disappeared when over 24,000 training samples were used.

The training window size had a clear effect on the model performance when training the models on the whole original dataset. The F1 scores for the classifiers CNN2 and CNN4, trained using randomly initialized model weights and by the transfer learning for all tested window sizes, are presented in Figure 6. The highest F1 scores were obtained for the 90 s window size.

The best accuracy scores of all classifiers trained with the full dataset for a window size with the highest average F1 score for each model are presented in Table 2. There were only minor differences in performance between the simpler CNN2 and the more complex CNN4 and between using transfer learning and training using randomly initialized model weights. For all the tested models, the clear optima were in the 60–120 s range.

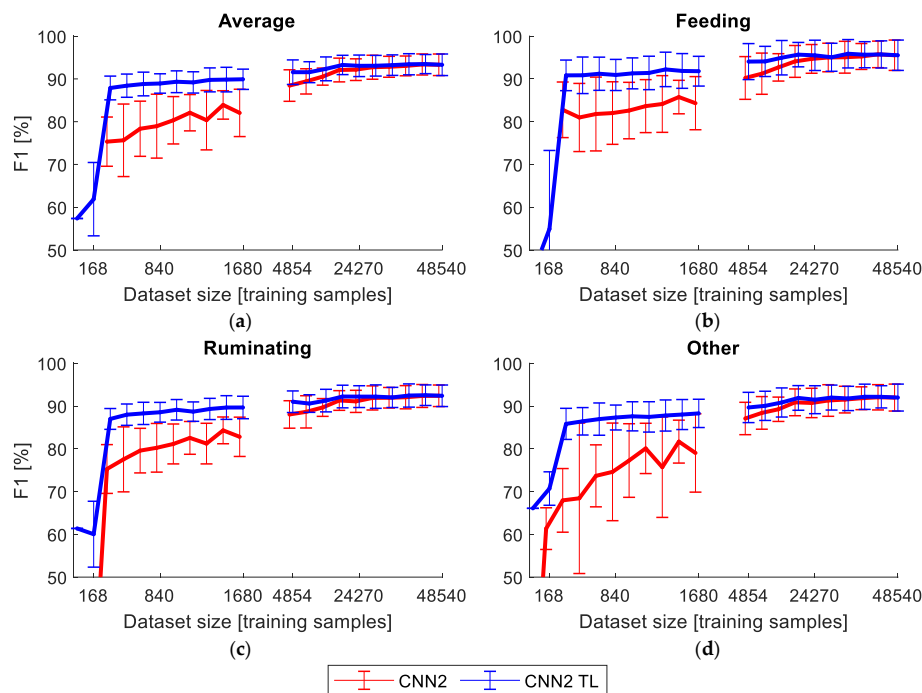


Figure 5. F1 score of the model CNN2 trained using randomly initialized model weights (CNN2) and by the transfer learning (CNN2 TL) for the window sizes of 60 s, depending on the training dataset size measured in training samples taken from the original dataset for the average F1 (a), feeding (b), rumination (c) and other behavior (d). The corresponding actual data after augmentation and balancing are: 336, ... ;3192, 13,878, ... ;132,762. The error bars represent the STD for 10-fold validation.

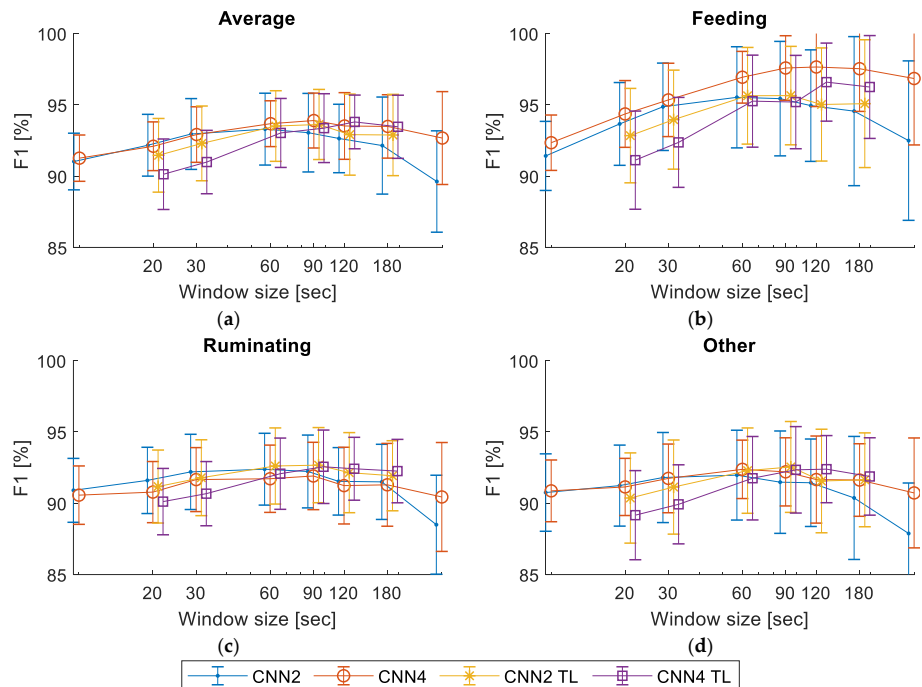


Figure 6. Performance of the tested models, CNN2 and CNN4, trained using randomly initialized model weights and by transfer learning (CNN2 TL and CNN4 TL), depending on the window size for the average F1 (a), feeding (b), rumination (c) and other behavior (d). The error bars represent the STD for 10-fold validation.

Table 2. Comparison of model performance trained by simple training and transfer learning. Precision, F1 and recall values are given averaged (mean \pm STD) for a 10-fold validation for optimal window sizes (WS), presented in Figure 6a.

	CNN2	CNN4	CNN2 TL	CNN4 TL
Precision	92.9 \pm 2.5	93.3 \pm 2.0	93.3 \pm 2.5	93.3 \pm 1.9
F1	93.3 \pm 2.5	93.9 \pm 1.9	93.6 \pm 2.4	93.8 \pm 1.8
Recall	94.2 \pm 1.7	94.3 \pm 1.5	94.5 \pm 2.5	94.4 \pm 1.4
WS (s)	60	90	90	120

The accuracy of the models trained on one dataset and validated by another was low: F1 = 57.3% for the model trained by the dataset published by [10] (the pretrained model before transfer learning), and F1 = 63.0% for the model trained by the dataset collected in this study and validated by dataset published by [10].

4. Discussion

Using the transfer learning technique, a relatively high (F1 = 87%) behavioral classification accuracy was reached with less than 500 training samples, a significantly better performance (12% higher) compared to the model initialized with random weights. An even higher classification performance was reached by using more training data. In this study, the average F1 score reached its maximum level with around 30,000 training samples, and the use of transfer learning was beneficial up to 24,000 training samples.

The differences in accuracy between the simpler CNN2 and the more complex CNN4 architectures and between the learning model using randomly initialized model weights and transfer learning were not significant when the full dataset was used. This may suggest that increasing the number of layers in an NN does not significantly increase the classifier accuracy for accelerometer data. Additionally, using transfer learning can improve the model performance when large amounts of training data cannot be collected.

The analysis of different training dataset sizes showed a constantly increasing accuracy when more training data was used. However, it also shows the limitation of the classifier, as increasing the dataset size to over 30% did not significantly increase the F1 score from the level of approximately 94%. Nevertheless, this effect occurred in this specific case for which the data were collected with specific sensors in the same environment for the same animals during a short period of time. The influence of the condition diversity on the accuracy of the models should be further studied.

The feeding behavior classification model with the best performance was CNN4, which had an average 93.9% F1 score for the 90 s optimal window size; this is close to the median among the values found in the literature review (Table A1). In practice, classification models are run on continuously measured data that are not split into windows containing data from only one behavior. A large window size can therefore create additional uncertainty in the classification accuracy because sample windows can contain mixed behavioral classes. The F1 score achieved in this study was high compared to the systems using an NN for cow behavior classification, such as those by [10] with F1 = 82%, [11] with F1 = 88.7% and [12] with F1 = 94.4%. Among the systems using machine learning with the performance reviewed in Table A1 were those by [16] with F1 = 93.3%, by [29] with F1 = 98.51%, and by [15] with a total accuracy of 98%. An exact comparison of the accuracy scores is impractical due to differences in the research conditions, such as the experimental environment, sensors and the amount and type of the collected data. However, the results in the mentioned studies can be applicable in actual conditions. Having the data openly available would be beneficial for comparison of the developed methods. Data collected in this study and other freely available datasets are listed at <https://github.com/Animal-Data-Inventory/PLFDataInventory> (accessed on 2 January 2023).

The low accuracy (F1 = 57.3% and F1 = 63.0%) of the models trained and validated by datasets achieved from different cows, different sensors and different reference data and

environments showed that both models were not applicable to other environments without additional fitting. However, the high accuracy of the transfer-learned models can mean that the basic patterns of the cow motions characterizing this class of problem were effectively learned by the convolutional layers of the models and used regardless the differences in the sensors and labeling methods.

The development of a machine-learning-based classifier includes a large number of elements and processes; hence, all model parameters could not be chosen optimally. Some of them were manually fine-tuned or adopted from previous studies. Among these parameters are ones related to the data collection (sampling frequency, number of measured axes, variability of animals and environments), data pre-processing (filter parameters, data fixing and augmentation methods) and NN architecture (number, type and size of the layers, number of filters, size of kernels, etc.).

During the experiments, it was found that augmentation simulating the sampling loss implemented by [25] did not improve the performance of the tested models. Due to frequent missing data intervals with a length of 10–20 samples, the data imputations performed by [29,30] were not effective for this system.

The information missing rate during the data transfer from the tags was about 50%, while it decreased when the number of tags was decreased. The main reason for this was connected with the large number of sent data messages, which led to hardware limitations [31]. The number of messages depends on the number of tags fitted to the cows and the sampling frequency, which was set to 25 Hz. However, the number of cows enrolled in this study was typical for a commercial barn, and this kind of measuring system should be able to perform in set conditions. An additional analysis of data redundancy must be conducted to find out to what extent the sampling frequency can be decreased in order to diminish the information loss or to increase the number of cows in the same compartment.

During the development and analysis of the classification model, a large number of assumptions were made to achieve practical results (e.g., the classification model and the estimation of a sufficient amount of the training data) in a realistic environment (e.g., low quality data, different sensors) with limited available data. The analysis of these assumptions must be performed in the future research. However, the results achieved in this study appear promising for practical applications.

In future work, additional uncertainties related to the training of classifiers should be studied. The limits of the applicability of trained models should be tested by application on datasets collected in different conditions and environments. To achieve this aim, additional data collection or the adoption of existing datasets is required. The influence of the data collected from a specific animal [32] on the classification accuracy of the entire herd should be also tested. The minimal amount of data transferred from the sensors should be evaluated by reducing the number of measured acceleration axes and the sampling frequency.

5. Conclusions

In this study, we developed a low-cost, open-source system for feeding behavior classification of dairy cows with an average F1 score equal to 93.9% and analyzed different methods and amounts of required training data. A dataset of approximately 20 cow × days for learning using randomly initialized model weights and approximately 10 cow × hours for the transfer learning were sufficient to achieve F1 = 90%. Despite a relatively high classification accuracy, additional research is needed to evaluate the applicability of the classifier to other environments and conditions.

Supplementary Materials: The structure of the CNN classifier models is available in the Supplementary Materials: https://github.com/cowbhav/CowBhave_BehaviorModel (accessed on 2 January 2023).

Author Contributions: Conceptualization, M.P. and V.B.; methodology, V.B.; software, M.P. and V.B.; validation, V.B. and M.M.; investigation, V.B.; resources, L.F.; data curation, V.B.; writing—original draft preparation, V.B.; writing—review and editing, M.P. and V.B.; visualization, V.B.; supervision, M.P.; project administration, M.P.; funding acquisition, M.P. and C.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the ICT-AGRI-2 ERA-NET project “CowBhave: a basic low-cost open-source automated monitoring system for discrimination of dairy cow behavioral activities” in Finland by the Ministry of Agriculture and Forestry.

Institutional Review Board Statement: The animal study protocol was approved by Animal Welfare body (Government decree 564/2013 22§) of Natural Resources Institute Finland. Project authorization was not needed as the experiment did not cause the animals a level of pain, suffering, distress or lasting harm equivalent to, or higher than, that caused by the introduction of a needle (2010/63/EU).

Informed Consent Statement: Not applicable.

Data Availability Statement: Bloch, V.; Frondelius, L.; Arcidiacono, C.; Mancino, M.; Pastell, M. 2022. Data from neck acceleration tags for training cow’s feeding behavior classifier. Zenodo, v1. <https://doi.org/10.5281/zenodo.6784671>.

Acknowledgments: The authors acknowledge CSC—IT Center for Science, Finland, for computational resources.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

To recognize the cow feeding behavior, different machine learning (ML) methods have been used in research studies, as listed in Table A1. A specific set of features was used in each study which was extracted from the acceleration samples in time and frequency domains. However, analyses of the feature types and their amount and physical interpretability have not been found.

Table A1. Review of methods for dairy cow behavior recognition based on acceleration sensors. Studies using NN-based methods are highlighted in bold.

Publication	Behavior Types	Interval Length, Sampling Rate	Number of Features	Method	Accuracy (F1)	Animals, Period, Barns
Achour, 2019 [33]	S, L, transition	3–10 s, 1–4 Hz	2	DT	99%	8, 0.25, 1
Arcidiacono, 2017 [16]	F, S	5 s, 4 Hz	1	DT	93.3%	5, 5 h, 1
Barwick, 2018 [34]	G, W, S, L	10 s, 12 Hz	12	Quadratic discriminant analysis		5, 2.5 h, 1
Benaissa, 2017 [35]	F, Ru, other activity	60 s, 10 Hz	9	DT, SVM	94.4%	10, 6 h, 1
Dutta, 2015 [36]	G, searching, W, Ru, Re, scratching	5 s, 10 Hz	9	probabilistic principal components analysis, fuzzy C means, self-organizing map	89%	
Eerdekens, 2020 (horses) [37]	S, W, trot, canter, roll, paw, flank watching	2.1 s, 25 Hz		CNN	97.84%	
Kaler, 2019 (sheep) [38]	W, S, L	7 s, 16 Hz	16	RF	80%	18, 1.6
Li, C., 2021 [12]	F, W, salting, Ru, Re	10 s, 25 Hz		CNN	94.4%	6, 6 h, 1
Pavlovic, 2021 [10]	F, Ru, Re	10 Hz, 90 s		CNN	82%	18, 6–18 d, 1
Pavlovic, 2022 [39]	F, Ru, Re	10 Hz, 90 s		Hidden Markov model, LDA, partial least squares discriminant analysis	83%	18, 6–18 d, 1
Peng, 2019 [11]	F, L, Ru, licking salt, moving, social licking and head butt	3.2–12.8 s, 20 Hz		RNN with LSTM, CNN	88.7%	6, ?
Rahman, 2018 [40]	G, S, Re, Ru	200 samples, 12 Hz	6	Majority voting, WEKA		?, ?

Table A1. Cont.

Publication	Behavior Types	Interval Length, Sampling Rate	Number of Features	Method	Accuracy (F1)	Animals, Period, Barns
Rayas-Amor, 2017 [3]	G, R	30 s	2	Linear regression	96.1(R2)	7, 9
Riaboff, 2020 [15]	G, W, Ru, Re	10 s,		Extreme boosting algorithm, Adaboost, SVM, RF	98% accuracy	86, 57 h, 4
Shen, 2019 [41]	F, Ru, O	256 samples, 5 Hz	30	K-nearest neighbor, SVM, PNN	92.4%	5, ?
Simanungkalit, 2021 [42]	Licking, F, S, L	10 s, 25 Hz	8	DT, RF, KNN, SVM	95–99% accuracy	4, 3.5 d
Tian, 2021 [28]	F, Ru, running, Re, head-shaking, drinking, W	?, 12.5 Hz	9	KNN, RF, KNN-RF fusion	99.34%	20, 3, 1
Vázquez Diosdado, 2015 [43]	F, S, L	300 s, 50 Hz		DT, SVM	91.7%	6, 1.5
Vázquez Diosdado, 2019 (sheep) [44]	W, S, L	7 s, 16 Hz	1	k-means, KNN	60.4%	26, 39
Walton, 2019 (sheep) [45]		5–7 s, 16–32 Hz	44	RF	91–97%	
Wang, 2018 [46]	F, L, S, W	5 s, 1 Hz		Adaptive boosting algorithm	75%	5, 25 h
Wang, 2020 [47]	Estrus	0.5–1.5 h, 1 Hz		KNN, back-propagation neural network, LDA, classification and regression tree	78.6–97.5%	12, 12 d
Williams, 2019 [48]	G, Re and W			13 ML algorithms	93%	40, 0.25

G—grazing; F—feeding; L—lying; O—other behaviors, excluding those already listed; Re—resting; Ru—rumination; S—standing; W—walking; SVM—support vector machine; DT—decision tree; LDA—linear discriminant analysis; KNN—K-nearest neighbor; ? (or empty)—data was not found in the publication.

References

- García, R.; Aguilar, J.; Toro, M.; Pinto, A.; Rodríguez, P. A systematic literature review on the use of machine learning in precision livestock farming. *Comput. Electron. Agric.* **2020**, *179*, 105826. [\[CrossRef\]](#)
- Borchers, M.R.; Chang, Y.M.; Proudfoot, K.L.; Wadsworth, B.A.; Stone, A.E.; Bewley, J.M. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *J. Dairy Sci.* **2017**, *100*, 5664–5674. [\[CrossRef\]](#)
- Rayas-Amor, A.A.; Morales-Almaráz, E.; Licona-Velázquez, G.; Vieyra-Alberto, R.; García-Martínez, A.; Martínez-García, C.G.; Cruz-Monterrosa, R.G.; Miranda-de la Lama, G.C. Triaxial accelerometers for recording grazing and ruminating time in dairy cows: An alternative to visual observations. *J. Vet. Behav.* **2017**, *20*, 102–108. [\[CrossRef\]](#)
- Grinter, L.N.; Campler, M.R.; Costa, J.H.C. Technical note: Validation of a behavior-monitoring collar's precision and accuracy to measure rumination, feeding, and resting time of lactating dairy cows. *J. Dairy Sci.* **2019**, *102*, 3487–3494. [\[CrossRef\]](#)
- Liu, L.S.; Ni, J.Q.; Zhao, R.Q.; Shen, M.X.; He, C.L.; Lu, M.Z. Design and test of a low-power acceleration sensor with Bluetooth Low Energy on ear tags for sow behaviour monitoring. *Biosyst. Eng.* **2018**, *176*, 162–171. [\[CrossRef\]](#)
- Arcidiacono, C.; Mancino, M.; Porto, S.M.C.; Bloch, V.; Pastell, M. IoT device-based data acquisition system with on-board computation of variables for cow behaviour recognition. *Comput. Electron. Agric.* **2021**, *191*, 106500. [\[CrossRef\]](#)
- Riaboff, L.; Shalloo, L.; Smeaton, A.F.; Couvreur, S.; Madouasse, A.; Keane, M.T. Predicting livestock behaviour using accelerometers: A systematic review of processing techniques for ruminant behaviour prediction from raw accelerometer data. *Comput. Electron. Agric.* **2022**, *192*, 106610. [\[CrossRef\]](#)
- Ferrari, A.; Micucci, D.; Mobilio, M.; Napoletano, P. Trends in human activity recognition using smartphones. *J. Reliab. Intell. Env.* **2021**, *7*, 189–213. [\[CrossRef\]](#)
- Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [\[CrossRef\]](#)
- Pavlovic, D.; Davison, C.; Hamilton, A.; Marko, O.; Atkinson, R.; Michie, C.; Crnojević, V.; Andonovic, I.; Bellekens, X.; Tachtatzis, C. Classification of Cattle Behaviours Using Neck-Mounted Accelerometer-Equipped Collars and Convolutional Neural Networks. *Sensors* **2021**, *21*, 4050. [\[CrossRef\]](#)
- Peng, Y.; Kondo, N.; Fujiura, T.; Suzuki, T.; Wulandari Yoshioka, H.; Itoyama, E. Classification of multiple cattle behavior patterns using a recurrent neural network with long short-term memory and inertial measurement units. *Comput. Electron. Agric.* **2019**, *157*, 247–253. [\[CrossRef\]](#)
- Li, C.; Tokgoz, K.K.; Fukawa, M.; Bartels, J.; Ohashi, T.; Takeda, K.; Ito, H. Data Augmentation for Inertial Sensor Data in CNNs for Cattle Behavior Classification. *IEEE Sens. Lett.* **2021**, *5*, 1–4. [\[CrossRef\]](#)
- Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [\[CrossRef\]](#)

14. Pastell, M.; Frondelius, L. A hidden Markov model to estimate the time dairy cows spend in feeder based on indoor positioning data. *Comput. Electron. Agric.* **2018**, *152*, 182–185. [[CrossRef](#)]
15. Riaboff, L.; Poggi, S.; Madouasse, A.; Couvreur, S.; Aubin, S.; Bédère, N.; Goumand, E.; Chauvin, A.; Plantier, G. Development of a methodological framework for a robust prediction of the main behaviours of dairy cows using a combination of machine learning algorithms on accelerometer data. *Comput. Electron. Agric.* **2020**, *169*, 105179. [[CrossRef](#)]
16. Arcidiacono, C.; Porto, S.M.; Mancino, M.; Cascone, G. Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data. *Comput. Electron. Agric.* **2017**, *134*, 124–134. [[CrossRef](#)]
17. Shahriar, M.S.; Smith, D.V.; Rahman, A.; Freeman, M.; Hills, J.; Rawnsley, R.P.; Henry, D.; Bishop-Hurley, G. Detecting heat events in dairy cows using accelerometers and unsupervised learning. *Comput. Electron. Agric.* **2016**, *128*, 20–26. [[CrossRef](#)]
18. WISDM HAR Dataset. Available online: <https://www.cis.fordham.edu/wisdm/dataset.php> (accessed on 2 January 2023).
19. Hu, J.; Zou, W.; Wang, J.; Pang, L. Minimum training sample size requirements for achieving high prediction accuracy with the BN model: A case study regarding seismic liquefaction. *Expert Syst. Appl.* **2021**, *185*, 115702. [[CrossRef](#)]
20. Kalouris, G.; Zacharaki, E.I.; Megalooikonomou, V. Improving CNN-based activity recognition by data augmentation and transfer learning. In Proceedings of the IEEE 17th International Conference on Industrial Informatics (INDIN), Helsinki, Finland, 22–25 July 2019. [[CrossRef](#)]
21. Eyobu, S.O.; Han, D.S. Feature Representation and Data Augmentation for Human Activity Classification Based on Wearable IMU Sensor Data Using a Deep LSTM Neural Network. *Sensors* **2018**, *18*, 2892. [[CrossRef](#)]
22. Oh, S.; Ashiquzzaman, A.; Lee, D.; Kim, Y.; Kim, J. Study on Human Activity Recognition Using Semi-Supervised Active Transfer Learning. *Sensors* **2021**, *21*, 2760. [[CrossRef](#)]
23. Li, F.; Shirahama, K.; Nisar, M.A.; Huang, X.; Grzegorzec, M. Deep Transfer Learning for Time Series Data Based on Sensor Modality Classification. *Sensors* **2020**, *20*, 4271. [[CrossRef](#)] [[PubMed](#)]
24. Bloch, V.; Pastell, M. Monitoring of Cow Location in a Barn by an Open-Source, Low-Cost, Low-Energy Bluetooth Tag System. *Sensors* **2020**, *20*, 3841. [[CrossRef](#)]
25. Hossain, T.; Ahad, M.A.R.; Inoue, S. A Method for Sensor-Based Activity Recognition in Missing Data Scenario. *Sensors* **2020**, *20*, 3811. [[CrossRef](#)]
26. Um, T.T.; Pfister, F.M.J.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data augmentation of wearable sensor data for parkinson’s disease monitoring using convolutional neural networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, New York, NY, USA, 13–17 November 2017. [[CrossRef](#)]
27. Brownlee, J. *Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python*, 1st ed.; Machine Learning Mastery; 2019; Available online: https://books.google.com.hk/books/about/Deep_Learning_for_Computer_Vision.html?id=DOamDwAAQBAJ&redir_esc=y (accessed on 2 January 2023).
28. Wang, Y.; Cang, S.; Yu, H. A survey on wearable sensor modality centred human activity recognition in health care. *Expert Syst. Appl.* **2019**, *137*, 167–190. [[CrossRef](#)]
29. Tian, F.; Wang, J.; Xiong, B.; Jiang, L.; Song, Z.; Li, F. Real-Time Behavioral Recognition in Dairy Cows Based on Geomagnetism and Acceleration Information. *IEEE Access* **2021**, *9*, 109497–109509. [[CrossRef](#)]
30. Weerakody, P.B.; Wong, K.W.; Wang, G.; Ela, W. A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing* **2021**, *441*, 161–178. [[CrossRef](#)]
31. Tosi, J.; Taffoni, F.; Santacatterina, M.; Sannino, R.; Formica, D. Performance Evaluation of Bluetooth Low Energy: A Systematic Review. *Sensors* **2017**, *17*, 2898. [[CrossRef](#)]
32. Wijekoon, A.; Wiratunga, N.; Sani, S.; Cooper, K. A knowledge-light approach to personalised and open-ended human activity recognition. *Knowl. -Based Syst.* **2020**, *192*, 105651. [[CrossRef](#)]
33. Achour, B.; Belkadi, M.; Aoudjit, R.; Laghrouche, M. Unsupervised automated monitoring of dairy cows’ behavior based on Inertial Measurement Unit attached to their back. *Comput. Electron. Agric.* **2019**, *167*, 105068. [[CrossRef](#)]
34. Barwick, J.; Lamb, D.W.; Dobos, R.; Welch, M.; Trotter, M. Categorising sheep activity using a tri-axial accelerometer. *Comput. Electron. Agric.* **2018**, *145*, 289–297. [[CrossRef](#)]
35. Benaissa, S.; Tuytens, F.A.M.; Plets, D.; de Pessemier, T.; Trogh, J.; Tanghe, E.; Martens, L.; Vandaele, L.; van Nuffel, A.; Joseph, W.; et al. On the use of on-cow accelerometers for the classification of behaviours in dairy barns. *Res. Vet. Sci.* **2019**, *125*, 425–433. [[CrossRef](#)]
36. Dutta, R.; Smith, D.; Rawnsley, R.; Bishop-Hurley, G.; Hills, J.; Timms, G.; Henry, D. Dynamic cattle behavioural classification using supervised ensemble classifiers. *Comput. Electron. Agric.* **2015**, *111*, 18–28. [[CrossRef](#)]
37. Eerdekens, A.; Deruyck, M.; Fontaine, J.; Martens, L.; De Poorter, E.; Joseph, W. Automatic equine activity detection by convolutional neural networks using accelerometer data. *Comput. Electron. Agric.* **2020**, *168*, 105139. [[CrossRef](#)]
38. Kaler, J.; Mitsch, J.; Vázquez-Diosdado, J.A.; Bollard, N.; Dottorini, T.; Ellis, K.A. Automated detection of lameness in sheep using machine learning approaches: Novel insights into behavioural differences among lame and non-lame sheep. *R. Soc. Open Sci.* **2020**, *7*, 190824. [[CrossRef](#)]
39. Pavlovic, D.; Czerkawski, M.; Davison, C.; Marko, O.; Michie, C.; Atkinson, R.; Crnojevic, V.; Andonovic, I.; Rajovic, V.; Kvascev, G.; et al. Behavioural Classification of Cattle Using Neck-Mounted Accelerometer-Equipped Collars. *Sensors* **2022**, *22*, 2323. [[CrossRef](#)]

40. Rahman, A.; Smith, D.; Little, B.; Ingham, A.; Greenwood, P.; Bishop-Hurley, G.J. Cattle behaviour classification from collar, halter, and ear tag sensors. *Inform. Process. Agric.* **2018**, *5*, 124–133. [[CrossRef](#)]
41. Shen, W.; Cheng, F.; Zhang, Y.; Wei, X.; Fu, Q.; Zhang, Y. Automatic recognition of ingestive-related behaviors of dairy cows based on triaxial acceleration. *Inf. Process. Agric.* **2020**, *7*, 427–443. [[CrossRef](#)]
42. Simanungkalit, G.; Barwick, J.; Cowley, F.; Dobos, R.; Hegarty, R. A Pilot Study Using Accelerometers to Characterise the Licking Behaviour of Pened Cattle at a Mineral Block Supplement. *Animals* **2021**, *11*, 1153. [[CrossRef](#)]
43. Vázquez Diosdado, J.A.; Barker, Z.E.; Hodges, H.R.; Amory, R.J.; Croft, D.P.; Bell, N.J.; Codling, E.A. Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry* **2015**, *3*, 1–14. [[CrossRef](#)]
44. Vázquez-Diosdado, J.A.; Paul, V.; Ellis, K.A.; Coates, D.; Loomba, R.; Kaler, J.A. Combined Offline and Online Algorithm for Real-Time and Long-Term Classification of Sheep Behaviour: Novel Approach for Precision Livestock Farming. *Sensors* **2019**, *19*, 3201. [[CrossRef](#)]
45. Walton, E.; Casey, C.; Mitsch, J.; Vázquez-Diosdado, J.A.; Yan, J.; Dottorini, T.; Ellis, K.A.; Winterlich, A.; Kaler, J. Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour. *R. Soc. Open Sci.* **2018**, *5*, 171442. [[CrossRef](#)] [[PubMed](#)]
46. Wang, J.; He, Z.; Zheng, G.; Gao, S.; Zhao, K. Development and validation of an ensemble classifier for real-time recognition of cow behavior patterns from accelerometer data and location data. *PLoS ONE* **2018**, *13*, e0203546. [[CrossRef](#)] [[PubMed](#)]
47. Wang, J.; Bell, M.; Liu, X.; Liu, G. Machine-Learning Techniques Can Enhance Dairy Cow Estrus Detection Using Location and Acceleration Data. *Animals* **2020**, *10*, 1160. [[CrossRef](#)]
48. Williams, M.L.; Wiliam, P.J.; Rose, M.T. Variable segmentation and ensemble classifiers for predicting dairy cow behaviour. *Biosyst. Eng.* **2019**, *178*, 156–167. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.