

Development of a recurrent neural networks-based calving prediction model using activity and behavioral data

Computers and Electronics in Agriculture

Keceli, Ali Seydi; Catal, Cagatay; Kaya, Aydin; Tekinerdogan, Bedir

<https://doi.org/10.1016/j.compag.2020.105285>

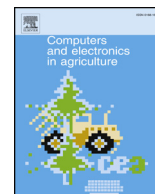
This article is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this article please contact openscience.library@wur.nl



Development of a recurrent neural networks-based calving prediction model using activity and behavioral data



Ali Seydi Keceli^{a,*}, Cagatay Catal^b, Aydin Kaya^c, Bedir Tekinerdogan^d

^a Department of Software Engineering, Cankaya University, Ankara, Turkey

^b Department of Computer Engineering, Bahcesehir University, Istanbul, Turkey

^c Department of Computer Engineering, Cankaya University, Ankara, Turkey

^d Information Technology Group, Wageningen University & Research, Wageningen, the Netherlands

ARTICLE INFO

Keywords:

Calving prediction
Recurrent neural networks
Machine learning
Precision dairy farming

ABSTRACT

Accurate prediction of calving time in dairy cattle is crucial for dairy herd management to reduce risks like dystocia and pain. Prediction of calving using traditional, manual observation such as observing breeding records and visual cues, however, is a complicated and error-prone task whereby even experts can fail to provide a proper prediction. Moreover, manual prediction does not scale for larger farms and becomes very soon time-consuming, inefficient, and costly. In this context, automated solutions are considered to be promising to provide both better and more efficient predictions, thereby supporting the health of the dairy cows and reducing the unnecessary overhead for farmers. Although the first automated solutions appear to have mainly focused on statistical solutions, currently, machine learning approaches are now increasingly being considered as a feasible and promising approach for accurate prediction of calving. In this context, the objective of this study is to develop machine learning-based prediction models that provide higher performance compared to the existing tools, methods, and techniques. This study shows that the calving of the cattle can be predicted by applying several behaviors of cattle, behavioral monitoring sensors, and machine learning models. Bi-directional Long Short-Term Memory (Bi-LSTM) method has been applied for the prediction of the calving day, and the RusBoosted Tree classifier has been used to predict the remaining 8 h before calving. The experimental results demonstrated that Bi-LSTM provides better performance compared to the LSTM algorithm in terms of classification accuracy, while the RusBoosted Tree algorithm predicts the remaining 8 h accurately before calving. Furthermore, Recurrent Neural Networks provide high performance for the prediction of calving day.

1. Introduction

Accurate prediction of calving time in dairy cattle is crucial for dairy herd management to reduce the risks like dystocia and pain (Schuenemann et al., 2011; Mainau and Manteca, 2011). There are indeed several signs which can be observed by farmers such as teat filling, vulva edema, udder distension, and behavioral changes to predict the calving time in dairy cows (Zehner et al., 2019) but routine controls to recognize these signs are mostly subjective, time-consuming, expensive, and require expertise (Ouellet et al., 2016), and even expert personnel may not accurately detect all calvings. Prediction of calving using traditional, manual observation such as observing breeding records and visual cues is a complicated and error-prone task whereby even experts can fail to provide a proper prediction. Moreover, manual prediction does not scale for larger farms and becomes very soon time-

consuming, inefficient, and costly. This problem seems to be an acknowledged one given the fact that the number of cattle per farm is increasing every year, while the number of dairy farms is decreasing (Ouellet et al., 2016; Raussi, 2003).

Apart from the impact on the daily tasks of farmers, parturition is a critical time for cows and calves (Borchers et al., 2017). Calf mortality, dystocia, and trauma to the cow and calf are some of the negative results during a difficult birth (Johanson and Berger, 2003; Mee, 2004). These outcomes adversely impact animal welfare and economy in farm (Borchers et al., 2017; Mee, 2004). Trauma can cause uterine infection and reduced milk production, which in turn causes extra costs and decreased fertility (Calcante et al., 2014; Dematawena and Berger, 1997). If the calving time can be predicted accurately, the environment, and cow-related negative impacts, which may cause calf injuries, can be reduced (Mee, 2004; Calcante et al., 2014). Also, the pain can be

* Corresponding author.

E-mail addresses: aliseydi@cankaya.edu.tr (A.S. Keceli), akaya@cankaya.edu.tr, aydinkaya83@gmail.com (A. Kaya), bedir.tekinerdogan@wur.nl (B. Tekinerdogan).

<https://doi.org/10.1016/j.compag.2020.105285>

Received 28 June 2019; Received in revised form 7 February 2020; Accepted 8 February 2020

Available online 20 February 2020

0168-1699/ © 2020 Elsevier B.V. All rights reserved.

mitigated, and reproductive performance can be improved (Mainau and Manteca, 2011; Bellows et al., 1988).

With the help of the dairy monitoring systems, calving dates can be predicted with objective methods instead of subjective approaches. Although there are some studies on automatic prediction of calving time, most of them are statistical-based studies, and machine learning-based prediction of calving time is still an open research area. Machine learning was addressed before in dairy science for mastitis and estrus, but there is a limited number of studies on calving prediction with machine learning. Several research studies have been performed for predicting the calving time by using several features such as body temperature changes (Aoki et al., 2005), ultrasound results (Wright et al., 1988), progesterone blood level (Streyl et al., 2011), and electrolyte concentrations (Bleul et al., 2006). These systems were not widely used because of the difficulties in execution, investment costs, and lack of expertise (Calcante et al., 2014). Therefore, nowadays, new systems using activity, lying, and ruminating behaviors are being developed (Borchers et al., 2017).

The exact moment of calving can be predicted recently with various protocols like ultrasound monitoring (Wright et al., 1988), measuring sulfate, and 17 β -estradiol levels in blood (Shah et al., 2007), observing levels of progesterone (Matsas et al., 1992) and evaluating the concentration of mammary gland secretions (Bleul et al., 2006). Some previous studies inspect body temperature before and after calving (Lammoglia et al., 1997; Burfeind et al., 2011). The body temperature can indeed be used for prediction, but current products in the market have not been validated yet. Although these methods have been developed, none of them have yet been adopted by the producers for commercial use. Visual observation of the cow is also a practical approach. A computer vision-based approach is proposed by Cangar et al. (2008). Cow motion trajectories, body orientation, and size are recorded and classified. Sumi et al. (2017) uses motion features and video sequences to detect calving. The features acquired from normal and calving situations are compared, and calving alert is provided in case of surpassing a given threshold.

A commonly used method in calving prediction is the monitoring of the lying activity (Robert et al., 2009; Bikker et al., 2014; Borchers et al., 2016; Mattachini et al., 2013). These values can be tracked by the technologies currently used in dairy farms. The other kind of features that can be used in calving prediction are feeding patterns and rumination (Huzzey et al., 2005; Schirmann et al., 2013). In the study proposed by Palombi et al. (2013), a sophisticated vaginal GSM device is used for monitoring of the calving process. This system also alarms before calving. The use of a GSM device makes this method expensive and non-feasible. The method proposed by Clark et al. (2015) combined rumination and lying features to predict calving. In addition to these features, the vaginal temperature is also used in another study proposed by Ouellet et al. (2016). Statistical methods are used in these studies to predict calving time.

A machine learning approach in calving prediction is proposed by Borchers et al. (2017). In their study, rumination, lying, and activity features are combined to predict the calving time. There is a two-phase prediction. The first phase is the prediction of calving day, and the second phase is the prediction of the last 8 h before the calving. Both of these predictions are made with well-known machine learning methods. Random forest, linear discriminant analysis (LDA), and neural network (NN) classifiers are trained for prediction. The best results are obtained with the NN classifier.

Another machine learning-based study is proposed by Rutten et al. (2017) to predict the calving hour. In this study, 24 h data before calving is used for prediction with a logistic regression classifier. Logistic regression classifier with rumination data is also used in Fadul et al. (2017) to predict 3 h before calving. Zehner et al. (2019) proposed a Naive Bayes classifier-based method to predict calving hour and day. A comparison table of our proposed method with the other studies in the literature is given in 1.

Table 1
Comparison with Related Studies.

Study	Method	Daily	Hourly	Time Series	Deep	Features	Machine Learning Methods
Borchers et al. (2017)	Classification	Yes	Yes, Bihourly prediction	No	No	Activity, Lying, Rumination	Random Forest, Neural Network, LDA
Rutten et al. (2017)	Classification	No	Yes	No	No	Rumination, Eating activity	Logistic Regression
Zehner et al. (2019)	Classification	Yes	Yes	No	No	Rumination, Jaw movements, Eating activity	Naive Bayes
Fadul et al. (2017)	Classification	No	Yes, Alerting 3 h left to Calving	No	No	Rumination, Jaw movements, Eating activity	Logistic Regression
Sumi et al. (2017)	Thresholding	Yes	Yes-Start of Calving	No	No	Visual monitoring	
Our study	Classification	Yes	Yes, Predicting 8 h left to calving	Yes	Yes	Activity, Lying, Rumination	Bi-LSTM, Rusboosted Tree

In this study, we aim to develop a fully automated approach based on machine learning algorithms and several behaviors of cattle such as lying, ruminating, and motion behaviors for predicting the calving time of cattle in dairy farms. More specifically, we focus on Long Short-Term Memory (LSTM) networks and Bi-directional Long Short-Term Memory (Bi-LSTM) networks which are a type of Recurrent Neural Networks (RNN) because they are considered as one of the most interesting deep learning types which have been applied for complex problems such as text generation, language translation, and image captioning (Brownlee, 2017). Compared to the Convolutional Neural Networks (CNN) and Multi-Layer Perceptron (MLP) algorithms, they differ because of the recurrent connections. RNN algorithms have been designed for sequence prediction problems, and in this study, we investigate their performance on the prediction of the calving day. In addition to the prediction of the day, we also perform experiments to predict the remaining 8 h before calving time. For this purpose, we evaluate the applicability of the RusBoosted Tree classifier algorithm.

As in Borchers et al.'s study (Borchers et al., 2017), we took a two-stages approach (i.e., identification of the day before calving and assessing behavior every 8 h) because as emphasized by Borchers et al. (2017), this two-stages approach requires fewer computations compared to the analysis of all 8-hourly data for all cows at each time. Otherwise, numerous computations would be required, and computational time would dramatically increase. We should also express that if the first stage fails to identify the day before calving, the second stage will not start. However, the performance of the first stage indicates that this potential flaw is not considered as a severe drawback in this system. Our experiments are performed on the dataset, which was generated based on the data of 53 Holstein cattle at the University of Kentucky Coldstream Dairy.

Two research questions are defined for each phase of the approach. Our research questions (RQs) are as follows:

- RQ-1: To what extent can calving prediction models benefit from Recurrent Neural Networks (specifically, LSTM and Bi-LSTM algorithms) for the prediction of the calving day of cattle?
- RQ-2: To what extent can ensemble learning classifiers (specifically, RusBoosted Tree) predict the remaining 8 h before calving time?

As shown in the RQs, we aim to build a two-stage approach that will first identify the calving day and then predict the remaining 8 h before calving time. The automated prediction of the calving time is vital to minimize efforts in manual observation and risks. Machine learning-based detection and prediction models are beneficial tools to support dairy farm management practices.

The remaining sections are organized as follows: Section 2 provides the methodology. Section 3 presents the experimental results. Section 4 provides discussion and validity threats. Section 5 presents the conclusion and future work.

2. Methodology

Two different methods are developed for the prediction of the remaining day and hour before calving. The workflows of daily and hourly prediction methods are given in Fig. 1. A Bi-LSTM based method that utilizes observation sequences is proposed for prediction of the calving day, and a binary RusBoosted Tree classifier is applied for the prediction of the 8 h before calving. For day prediction, observations of the last two weeks before calving are used. The day before the calving is taken as a class label, which should be estimated, and the observations before this day are transformed into an input sequence. For example, to predict the day before calving, which is labeled as 1, all daily observations before this day are used.

The input sequences are split into testing and training sets for experimentation. Splitting (a.k.a., hold-out approach) is one of the standard methods for estimating the model performance in machine

learning. The dataset is randomly divided into disjoint training and testing sets. Randomly selected 80% of the data are used as the training set, and the remaining 20% of the data are used as the testing set in experiments. A Bi-LSTM classifier is trained with the training set, and the performance of this classifier is measured with testing data. The observations of the last 24 h are labeled in binary and classified with a RusBoosted Tree Classifier. The hours closer than 8 h are marked as 1, and the others are labeled as 0 to reduce the complexity of the classification problem. In the following subsections, we discuss each step in more detail.

2.1. Predicting calving day

Traditional neural networks make predictions based on discrete or single observations. For making predictions from a set of observation series, Recurrent Neural Networks (RNN) models are developed (Hochreiter and Schmidhuber, 1997). Automated image captioning, text translation, and handwriting generation are some of the popular applications of RNNs (Brownlee, 2017). It is assumed that all inputs (and outputs) in a conventional neural network are independent of each other. RNNs are called recursive because they perform the same task for each element of a sequence, and the output depends on previous observations. Unlike feedforward neural networks, RNNs can use their input memory to process arbitrary sequences of inputs. RNNs collect information about the observed values of the sequence and memorize this information. RNNs can theoretically be used for very long sequences, but in practice, they are limited to only a few steps back. Bi-LSTM and LSTM are some examples of RNN networks. The LSTMs and Bi-LSTMs have recurrent connections to use activations from the previous step to formulate the output. The LSTM units, which are called memory cell or memory blocks, are composed of weights and gates. Memory cells are the basic unit of the block for both Bi-LSTM and LSTM networks. These blocks are recurrently connected. The weights of these units are parameters for input, output, and internal state and computed through input sequences. The gates are the weighted functions to gather information from the sequence. There are three types of gates. These are input, output, and forget gates. Input gates select the input values to update the internal state. The output gates produce output according to internal state and input values. The forget gates decide which information will be cut out. A sample LSTM cell is shown in Fig. 2.

The main limitation of the LSTM units is the memory. The memory units need as much memory bandwidth as the number of operations and parameters. For problems like autoregression, it is better to use a windowed MLP (Multi-Layer Perceptron). Bi-LSTMs can be considered as the extended version of LSTM networks. Unlike the LSTM network, Bi-LSTMs process the input sequences in both forward and backward directions. In Bi-LSTM, there are duplicated recurrent layers. The first layer takes the input sequence as-is and gives the reversed input sequence to the second layer. An example Bi-LSTM is shown in Fig. 3. The X_s are serial inputs, LSTMs are the long short term memory units, and the H is output. The gates are symbolized with σ .

In our study, a Bi-LSTM network has been developed. The proposed network has an input layer with 9 input points, 1 Bi-LSTM layer with 500 hidden units, a fully-connected output layer, a soft-max layer, and finally, a classification layer. Adam optimizer is selected during experiments. The sgdm optimizer is also tested during the experiments, but no significant results obtained with it and the accuracy of the network trained with SDGM is low. The Bi-LSTM network is trained with the following parameters: gradient threshold = 1, initial learning rate = $1e-3$, the maximum number of epochs = 900, and Adam solver is used.

Input layers are the layers that read the observation series as inputs and pass them to the LSTM layers. The LSTM layers are estimated based on calculations over the sequence. Neurons in a fully-connected layer have full links to all the activations in the previous layer. For each class in the softmax layer, a softmax function is applied to the probability

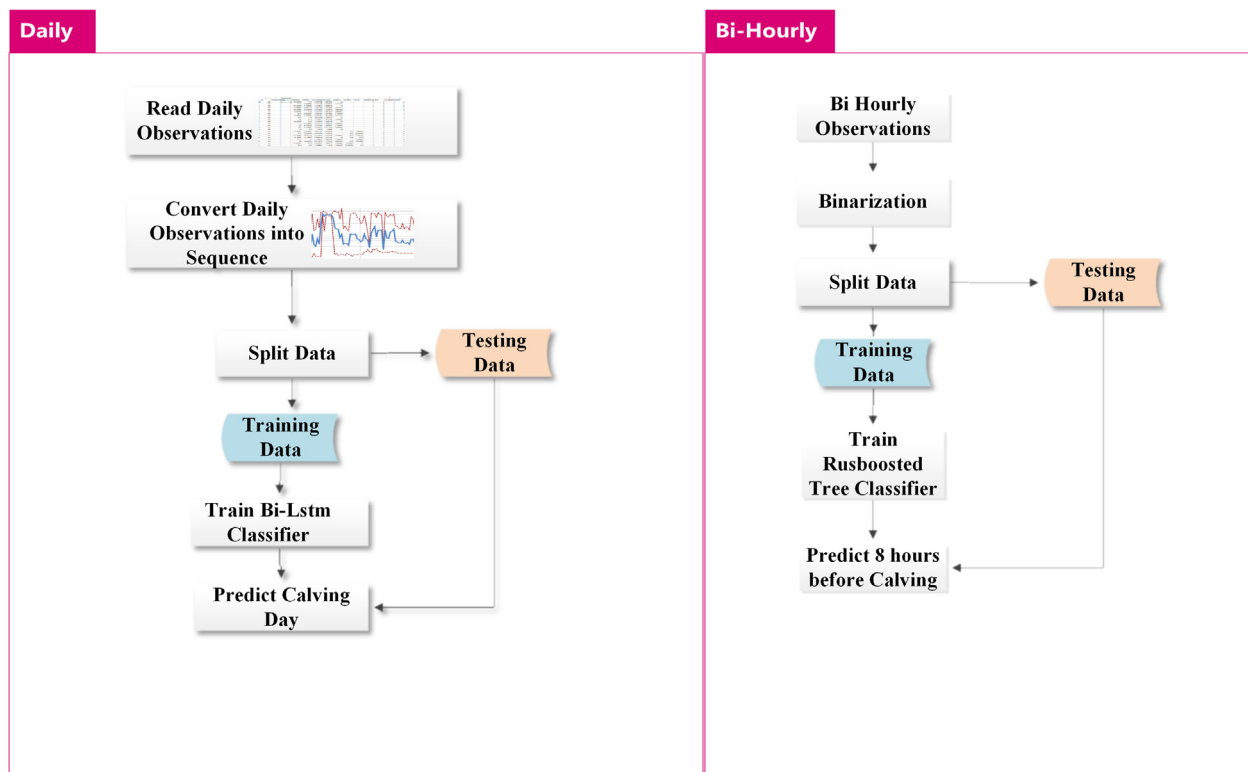


Fig. 1. General Flow of Proposed Method for daily and bi-hourly prediction.

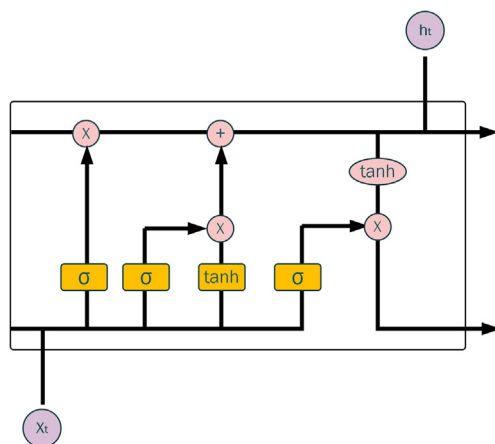


Fig. 2. A sample LSTM cell (Lstm unit, 2019).

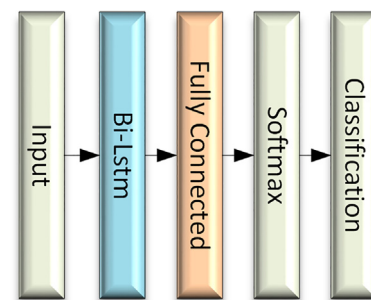


Fig. 4. Day Prediction Network.

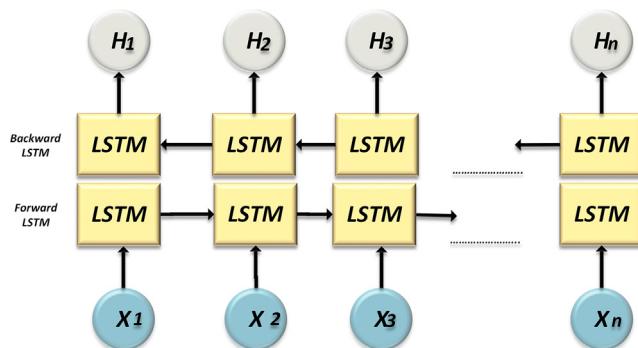


Fig. 3. A sample Bi-LSTM network (Keceli and Kaya, 2018).

values, and all the entries are completed to 1. The architecture of the developed Bi-LSTM network is shown in Fig. 4.

For predicting the number of days before calving, the measurements up to 14 days are used as sequential data. In the dataset generated by Borchers et al. (2017), there is day by day measurements before the calving for each cow. To predict a particular day, the measurements before that day are used. The features belonging to different days are organized as a sequence according to the day order and given to the Bi-LSTM network as input. For some of the samples in the dataset, the whole data before calving is not available, and there are missing days. The RNN classifiers are capable of classifying the series with different sizes, and therefore, these series are handled accordingly.

2.2. Predicting calving hour

For hourly prediction, a RusBoosted Tree classifier is applied. RusBoosted Tree classifier is an ensemble method that utilizes the RusBoost data balancing algorithm (Seiffert et al., 2010). It is a hybrid sampling/boosting algorithm based on SMOTEBoost (Chawla et al., 2003). The main benefit of the RUSBoost compared to SMOTEBoost is its simplicity. This method also produces a small number of samples

during the balancing process. The RUSBoosted tree is a collection of weak classifiers. These are classifiers that are trained with the random undersampled balanced data and have some clue on how to predict the right labels. The training data is first balanced with the RUSBoost algorithm in RUSBoosted trees. Then, the balanced samples are used in the training of the individual tree classifiers. The main difference between the RusBoost and the SMOTEBoost is performing random undersampling (RUS) in each boosting iteration, while SMOTEBoost gives more examples of the minority class. RusBoost algorithm is a combination of random undersampling and boosting. The method randomly deletes data from the majority class until the desired class distribution is achieved. Then, boosting can be performed with re-weighting and resampling. The RusBoost has no intelligent approach to remove samples. Instead, RusBoost removes samples randomly. In general, RusBoost slightly outperforms SMOTE when looking at the recall of the minority classes. In the first step of the RusBoost algorithm, a weight is given to each sample. In step 2, random oversampling is applied to balance the class ratio. As a result, new weight distribution is acquired. The new dataset is passed to the base learner (Decision Tree for RusBoosted tree). Finally, the new weights are re-calculated to use in the following iteration. This process is repeated until reaching the maximum number of iteration.

The actual hour prediction of calving is difficult due to a large number of labels. If we consider each hour as a label, there will be a high misclassification between close hours. Some observations between sequential hours are very similar, and this case is difficult for a predictive model to discriminate. As mentioned before, in order to reduce the number of classes, observations are labeled in binary, and the ones closer than 8 h are labeled as 1, and the others are labeled as 0. This approach is also used in reference (Borchers et al., 2017). When we tried to predict each hour separately, the overall correct classification ratio is very low. The labor and delivery process lasts less than 8 h before calving (The cow-calf manager, 2019). So the prediction of the last 8 h before calving is essential. Also, the study proposed in Borchers et al. (2017), best predictions are obtained for 8 h period before calving. To improve the classification ratio, we transformed the task into a binary classification problem. The observations recorded between 1–8 h before calving are labeled as 1, and the remaining hourly observations are labeled as 0. There is a tuple for each hour before calving. The last 24 h are used in our experiments.

We also tested Bi-LSTM and LSTM for hourly prediction, but low classification accuracies are obtained. The prediction accuracy is below 40% for both Bi-LSTM and LSTM classifiers. In our experiments, the best results are obtained with the binary RusBoosted Tree classifier for hourly prediction.

3. Experimental results

3.1. Dataset

In this study, the dataset collected by Borchers et al. (2017) which is publicly available through the following github repository, has been used: <https://github.com/Mrborchers/Machine-learning-based-calving-prediction-from-activity-lying-and-rumination-behaviors>. Data were collected from the University of Kentucky Coldstream Milk Plant with 20 primiparoses and 33 multiple rectum Holstein milk cattle. Cows were transported to dry cow plants 30 days before the calving date and placed in a package with 9.15–21.34-m Chip beds providing continuous access to 3.64 ha pasture. Two technological devices (HR label and IceQube) were placed on each cow 28 days before the predicted calves. After the calf, the data were reduced to include only two weeks before the calf was taken. HR label (SCR Engineers Ltd.) was placed on the left side of the neck, and the neck activity and rumination data were collected automatically during two hours periods using a 3-axis accelerometer and a microprocessor. The number of the left rear leg, the number of automatic steps, the time spent lying, the time spent lying,

the time spent standing, the number of standing passes, and the use of the 3-axis accelerometer were collected in 15 min periods. Third-party technological devices, namely the HR tag (Schirmann et al., 2009) and IceQube (McGowan et al., 1999; Borchers et al., 2016) were already validated in previous studies. Cows in the dry pencil were monitored every 3 h for signs of a calf. After the first sign was detected, every 15 min a unit cow was followed. For each calf, the calf date, cow parity, calving beginning time, and the time being outside the cow were recorded.

3.2. Results

Sensitivity, specificity, positive (TP, FP), and negative predictive values (TN, FN) are observed during the experiments to measure the effectiveness of the proposed models. Sensitivity is the proportion of actual positives that are correctly classified, and the specificity is the portion of actual negatives that are correctly detected. The formulation of sensitivity and specificity are provided in Eq. 1, 2. In equations, the expansions of the acronyms are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP is the number of correctly identified samples. FP is the number of incorrectly identified samples. TN is the number of correctly rejected samples.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

The positive and negative predictive values (PPV, NPV) are the proportions of positive and negative results in classification tasks. The equations are given as follows:

$$\text{PPV} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{NPV} = \frac{TN}{TN + FN} \quad (4)$$

In Table 2, Fig. 5 and Fig. 6, sensitivity, specificity, positive predictive values, and negative predictive values for LSTM and Bi-LSTM method for daily calving prediction are presented. Two classification methods performed better in specifying the positive class than specifying the negative class. According to the daily results, both of the methods were unsuccessful on day six, nine, and eleven. For both models, the sensitivity values of all classes (days) are admissible. This means that our models are reliable on the detection of actual class or calving day. A similar situation is observed for the specificity metric as well. This means that the day which is aimed to be classified and the other days

Table 2

Experimental results of LSTM and Bi-LSTM methods for daily calving prediction.

Days	LSTM				Bi-LSTM			
	Sens.	Spec.	PPV	NPV	Sens.	Spec.	PPV	NPV
1	0.86	0.98	0.75	0.99	1.00	1.00	1.00	1.00
2	0.43	0.98	0.60	0.95	1.00	0.99	0.88	1.00
3	0.50	0.98	0.67	0.95	0.71	0.99	0.83	0.98
4	0.67	0.94	0.29	0.99	0.75	0.99	0.86	0.98
5	0.71	0.99	0.83	0.98	0.71	0.99	0.83	0.98
6	0.40	0.99	0.67	0.97	0.83	0.94	0.50	0.99
7	1.00	0.95	0.50	1.00	1.00	1.00	1.00	1.00
8	0.83	0.99	0.83	0.99	0.90	1.00	1.00	0.99
9	0.29	0.99	0.67	0.94	0.67	0.99	0.80	0.98
10	0.75	0.98	0.60	0.99	0.86	0.99	0.86	0.99
11	0.80	0.99	0.80	0.99	0.50	0.99	0.50	0.99
12	0.91	0.99	0.91	0.99	0.80	0.99	0.80	0.99
13	1.00	0.98	0.80	1.00	0.88	0.99	0.88	0.99
14	0.88	1.00	1.00	0.99	0.86	0.99	0.86	0.99

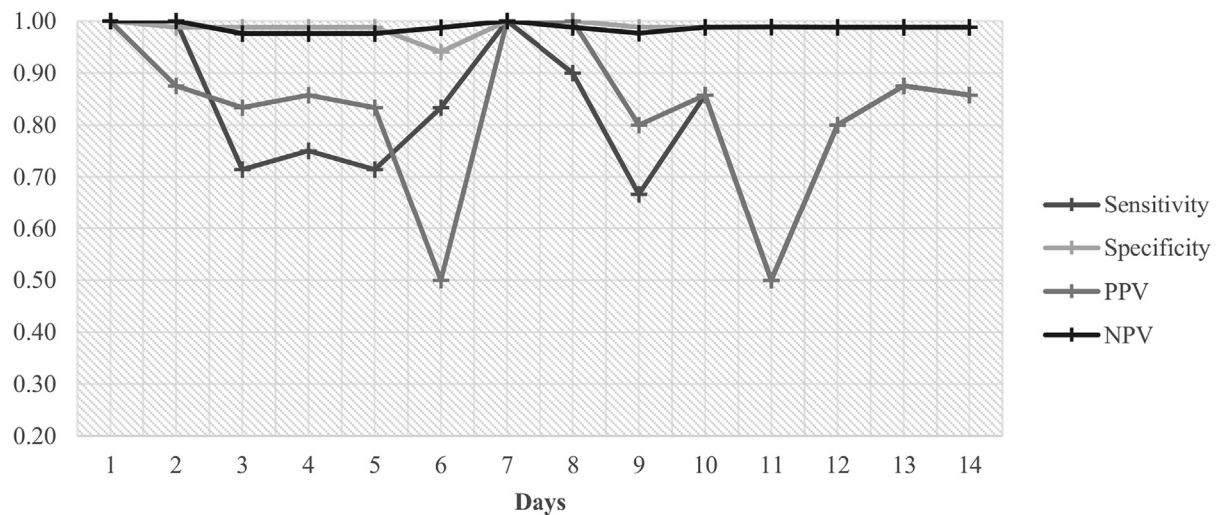


Fig. 5. Sensitivity, specificity, positive predictive values, and negative predictive values for Bi-LSTM method.

are successfully assigned to different classes. PPV can be evaluated as class-based accuracy. As shown in Fig. 5, for the first five days before calving, PPV values are high. The prediction of these days is performed better than the remaining days. The PPV values for the remaining days are unstable, both high and low PPV values can be seen for different days. As a result, the results of Bi-LSTM are acceptable, especially in the last five days of the calving.

According to figures for daily prediction, there is a high accuracy on the prediction of the day before calving. The prediction performance for the days before the first five days are not reliable. In Table 3, average scores for Bi-LSTM and LSTM are presented. Bi-LSTM and LSTM provided a classification accuracy of 83.34%, and 72.22%, respectively. The methods performed better on specificity than sensitivity. Bi-LSTM provided 81.91% mean sensitivity and 98.72% mean specificity. LSTM provided 71.57% mean sensitivity and 97.88% mean specificity. While both methods perform similarly in terms of specificity, Bi-LSTM performs better on sensitivity scores. Negative predictive values of both methods were very similar, yet Bi-LSTM performed better on the positive predictive value metric.

In our experiments, we also tested Bi-LSTM and LSTM with single inputs rather than series. Daily observations are used as inputs for RNN networks. The average classification accuracy for both LSTM and Bi-LSTM networks is lower compared with the series input the results. By utilizing the measurements from the previous days, a serious prediction

Table 3

Average scores for daily predictions.

	CA	Sens	Spec	PPV	NPV
LSTM	0.72	0.72	0.98	0.71	0.98
Bi-LSTM	0.83	0.82	0.99	0.83	0.99

Table 4

Results with Classical Methods.

	Sens	Spec	PPV	NPV
RF Borchers et al. (2017)	0.25	0.89	0.16	0.93
LDA Borchers et al. (2017)	0.75	0.93	0.50	0.97
K-NN	0.22	0.12	0.18	0.97
SVM	0.23	0.37	0.38	0.90
LSTM	0.72	0.98	0.71	0.98
Bi-LSTM	0.82	0.99	0.83	0.99

performance improvement has gained. The results show that using knowledge from earlier days is more descriptive for daily prediction. The results obtained from the classical machine learning methods like Random Forest, Linear Discriminant Analysis, SVM, and K-NN is presented in Table 4 with the results obtained from our RNN models for the day before the calving. As seen in these results, RNN models have much

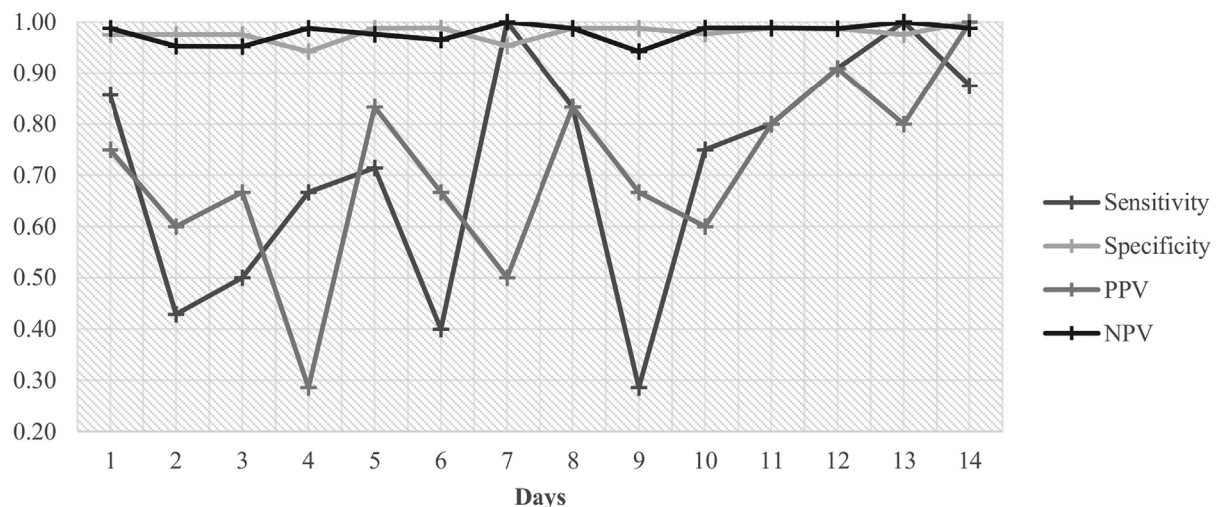


Fig. 6. Sensitivity, specificity, positive predictive values, and negative predictive values for LSTM method.

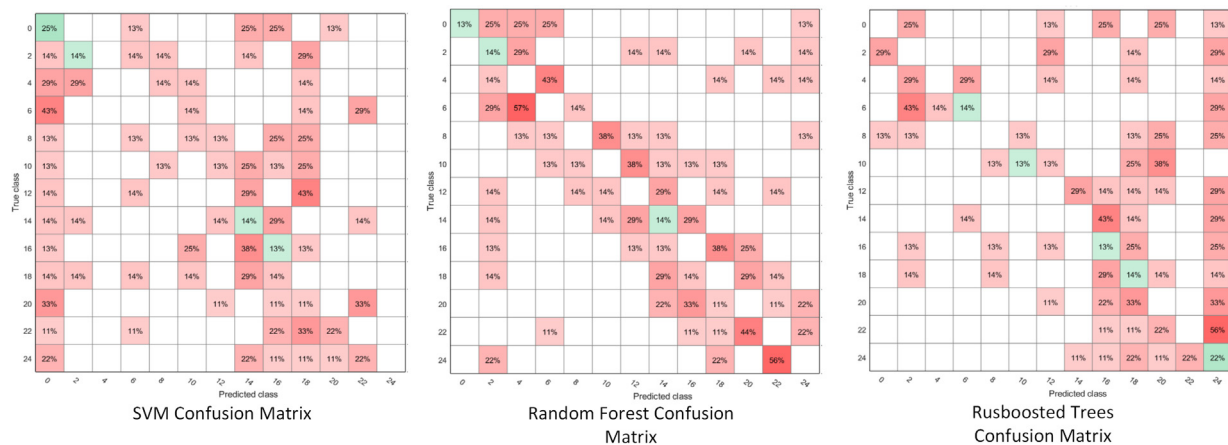


Fig. 7. Confusion matrices obtained from multi-class classification.

higher sensitivity and specificity values.

The data used for hourly prediction were recorded similar to the method proposed by Jensen (2012). For every cow, a bi-hourly measurement is made and subtracted from average values of the other days to determine the deviation from baseline values. The procedure is applied for all the features acquired.

Calving prediction has critical importance due to allowing the provision of expert assistance when necessary for livestock profitability and animal welfare. The actual hour prediction of calving is difficult due to a large number of labels. If we consider each hour as a label, the data provided between hours is very similar, and in this case, it is challenging for a predictive model to discriminate. The confusion matrices obtained when each bi-hour is considered as a separate class are shown in Fig. 7. The discrimination capacity of the models is too low, and there is a high ratio of misclassification. As seen from the confusion matrices, the classification accuracies of the multi-class models are lower than 30%. Especially the true prediction ratio of the last 8 h is too small, which is also critical for expert help. For coping with this situation and train more accurate models, the number of classes is reduced. The confusion matrix of the binary classification is presented in Fig. 8. As seen in Fig. 8, a more accurate model is trained with binary data.

For reducing the number of classes, the observations are labeled in binary. The ones closer than 8 h are marked as 1, and the others are labeled as 0. RUSBoosted Tree method was applied on hourly data. The method achieved a mean of 84.16% classification accuracy, 80.51% sensitivity and specificity, and 85.74% negative and positive predictive value. Results based on hour data are presented in Table 5.

4. Discussion

The background and related work section showed that Bi-LSTM and RusBoosted Tree classifiers had not been investigated for calving prediction yet. We investigated the applicability of RNN algorithms on the prediction of calving day and the ensemble learning classifier (RusBoosted Tree) for predicting the remaining 8 h before calving time. Our research questions were presented in Section 1, and in this section, we discuss our research findings regarding these research questions.

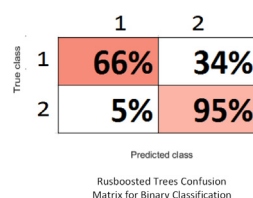


Fig. 8. Confusion matrix obtained from binary classification.

Table 5

Hourly prediction results.

	Sens	Spec	PPV	NPV
< 8 h	0.66	0.95	0.89	0.82
> 8 h	0.95	0.66	0.82	0.89

Our experimental results demonstrated that Bi-LSTM based prediction model provides better performance than the LSTM-based model in terms of classification accuracy. Also, the RusBoosted Tree algorithm-based prediction model accurately predicts the remaining 8 h before calving time. This study shows the power of Recurrent Neural Networks algorithms for calving prediction problem. Since this article is an experimental study, we should address the potential threats to validity in this section. For internal validity, we mainly evaluated the performance of two RNN algorithms (LSTM and Bi-LSTM), but new studies might investigate the performance of other LSTM-based algorithms. Although we selected the most promising and relevant RNN algorithms, there might be other LSTM-based algorithms that can improve the performance presented in this study.

Regarding the external validity, we performed our experiments on a large dataset, including 53 Holstein cattle, but the performance of the models might change on a different dataset having more or fewer data. Since we reached only to this public dataset, we performed our experiments only on this dataset. When more and more datasets become publicly available, we can implement new experiments by using our models presented in this study. Regarding the construct validity, we followed the experimental design approach of the researchers who generated this dataset. Regarding the conclusion validity, the original dataset has been divided into training and test datasets, and evaluation was performed on the testing sets in other machine learning studies. The results for the day before calving is presented with the best results obtained in Borchers et al. (2017) at Table 6. Although LSTM results are slightly lower, Bi-LSTM results are comparable and sometimes better.

Deep learnings superiority comes from its flexibility to learning to represent data as a nested hierarchy of concepts and extracting more abstract representations of the data. A deep learning architecture can

Table 6

The comparison of metrics obtained for the day before calving.

Metric	Bi-LSTM	LSTM	Ref. Borchers et al. (2017)
Sens	100	86	100
Spec	100	98	86.8
PPV	100	75	40.0
NPV	100	99	100

learn both low and high levels of features in its different layers. Another advantage of deep learning is related to the amount of data used in training. In traditional machine learning methods, classifiers performance is stuck after a certain amount of data. This restriction makes deep-based methods more successful compared to the traditional ones. Ensemble models are a combination of two or more single models. The idea behind ensemble methods is improving or boosting performance by combining different estimators. In other words, it is the process of pooling predictions of multiple base models. Empirical and theoretical studies show that model combination boosts predictive accuracy more often than the single models (Finlay, 2011; Paleologo et al., 2010).

5. Conclusion

Accurate and automated prediction of calving time of cattle in dairy farms has many advantages such as minimizing the daily tasks of farmers, mitigating the negative results of difficult births, and increasing the potential fertility in the farm. Instead of manually observing each cattle, it is possible to develop automated prediction systems by using different sensors that can determine the behaviors of cattle. In this study, we applied the RNN algorithms for the prediction of the calving day using a dataset on the motion, lying, and ruminating behaviors of cattle. We demonstrated that Bi-LSTM based prediction algorithm provides promising results in terms of classification accuracy. Besides, we showed that the RusBoosted Tree classifier, which is an ensemble learning algorithm accurately predicts the remaining 8 h before calving time. In our future work, we aim to apply the resulting models on new datasets when they become publicly available. Further, we plan to use novel RNN-based algorithms for enhancing even further the accuracy of the prediction of calving day.

Declaration of Competing Interest

The authors declare no conflict of interest.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.compag.2020.105285>.

References

- Aoki, M., Kimura, K., Suzuki, O., 2005. Predicting time of parturition from changing vaginal temperature measured by data-logging apparatus in beef cows with twin fetuses. *Anim. Reprod. Sci.* 86 (1–2), 1–12.
- Bellows, R., Short, R., Staigmiller, R., Milmine, W., 1988. Effects of induced parturition and early obstetrical assistance in beef cattle. *J. Anim. Sci.* 66 (5), 1073–1080.
- Bikker, J.P., van Laar, H., Rump, P., Doorenbos, J., van Meurs, K., Griffioen, G.M., Dijkstra, J., 2014. Technical note: evaluation of an ear-attached movement sensor to record cow feeding behavior and activity. *J. Dairy Sci.* 97 (5), 2974–2979. <https://doi.org/10.3168/jds.2013-7560>.
- Bleul, U., Spirig, S., Hässig, M., Kähn, W., 2006. Electrolytes in bovine prepartum mammary secretions and their usefulness for predicting parturition. *J. Dairy Sci.* 89 (8), 3059–3065.
- Bleul, U., Spirig, S., Hassig, M., Kahn, W., 2006. Electrolytes in bovine prepartum mammary secretions and their usefulness for predicting parturition. *J. Dairy Sci.* 89 (8), 3059–3065. [https://doi.org/10.3168/jds.S0022-0302\(06\)72580-2](https://doi.org/10.3168/jds.S0022-0302(06)72580-2).
- Borchers, M.R., Chang, Y.M., Tsai, I.C., Wadsworth, B.A., Bewley, J.M., 2016. A validation of technologies monitoring dairy cow feeding, ruminating, and lying behaviors. *J. Dairy Sci.* 99 (9), 7458–7466. <https://doi.org/10.3168/jds.2015-10843>.
- Borchers, M.R., Chang, Y.M., Proudfoot, K.L., Wadsworth, B.A., Stone, A.E., Bewley, J.M., 2017. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *J. Dairy Sci.* 100 (7), 5664–5674. <https://doi.org/10.3168/jds.2016-11526>.
- Brownlee, J., 2017. Long Short-term Memory Networks with Python: Develop Sequence Prediction Models with Deep Learning. Jason Brownlee.
- Burfeind, O., Suthar, V.S., Voigtsberger, R., Bonk, S., Heuwieser, W., 2011. Validity of prepartum changes in vaginal and rectal temperature to predict calving in dairy cows. *J. Dairy Sci.* 94 (10), 5053–5061. <https://doi.org/10.3168/jds.2011-4484>.
- Calcante, A., Tangorra, F.M., Marchesi, G., Lazzari, M., 2014. A gps/gsm based birth alarm system for grazing cows. *Comput. Electron. Agric.* 100, 123–130.
- Cangar, O., Leroy, T., Guarino, M., Vranken, E., Fallon, R., Lenehan, J., Mee, J., Berckmans, D., 2008. Automatic real-time monitoring of locomotion and posture behaviour of pregnant cows prior to calving using online image analysis. *Comput. Electron. Agric.* 64 (1), 53–60. <https://doi.org/10.1016/j.compag.2008.05.014>.
- Chawla, N.V., Lazarevic, A., Hall, L.O., Bowyer, K.W., 2003. Smoteboost: Improving prediction of the minority class in boosting. In: *Knowledge Discovery in Databases: Pkdd 2003*, Proceedings 2838, pp. 107–119.
- Clark, C.E.F., Lyons, N.A., Millapan, L., Talukder, S., Cronin, G.M., Kerrisk, K.L., Garcia, S.C., 2015. Rumination and activity levels as predictors of calving for dairy cows. *Animal* 9 (4), 691–695. <https://doi.org/10.1017/S1751731114003127>.
- Dematawena, C., Berger, P., 1997. Effect of dystocia on yield, fertility, and cow losses and an economic evaluation of dystocia scores for holsteins. *J. Dairy Sci.* 80 (4), 754–761.
- Fadul, M., Bogdahn, C., Alsaad, M., Hüsler, J., Starke, A., Steiner, A., Hirsbrunner, G., 2017. Prediction of calving time in dairy cattle. *Anim. Reprod. Sci.* 187, 37–46.
- Finlay, S., 2011. Multiple classifier architectures and their application to credit risk assessment. *Eur. J. Oper. Res.* 210 (2), 368–378.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Huzzey, J.M., von Keyserlingk, M.A.G., Weary, D.M., 2005. Changes in feeding, drinking, and standing behavior of dairy cows during the transition period. *J. Dairy Sci.* 88 (7), 2454–2461. [https://doi.org/10.3168/jds.S0022-0302\(05\)72923-4](https://doi.org/10.3168/jds.S0022-0302(05)72923-4).
- Jensen, M.B., 2012. Behaviour around the time of calving in dairy cows. *Appl. Anim. Behav. Sci.* 139 (3–4), 195–202.
- Johanson, J., Berger, P., 2003. Birth weight as a predictor of calving ease and perinatal mortality in holstein cattle. *J. Dairy Sci.* 86 (11), 3745–3755.
- Keceli, A.S., Kaya, A., 2018. Classification of violent activities with optical flow image and bi-lstm. *Eur. J. Sci. Technol.* (14), 204–208.
- Lammoglia, M.A., Bellows, R.A., Short, R.E., Bellows, S.E., Bighorn, E.G., Stevenson, J.S., Randel, R.D., 1997. Body temperature and endocrine interactions before and after calving in beef cows. *J. Anim. Sci.* 75 (9), 2526–2534.
- Lstm unit, <<https://www.oreilly.com/ideas/introduction-to-lstms-with-tensorflow>>, accessed: 2019-08-10.
- Mainau, E., Manteca, X., 2011. Pain and discomfort caused by parturition in cows and sows. *Appl. Anim. Behav. Sci.* 135 (3), 241–251.
- Mainau, E., Manteca, X., 2011. Pain and discomfort caused by parturition in cows and sows. *Appl. Anim. Behav. Sci.* 135 (3), 241–251. <https://doi.org/10.1016/j.applanim.2011.10.020>.
- Matsas, D.J., Nebel, R.L., Pelzer, K.D., 1992. Evaluation of an on-farm blood progesterone test for predicting the day of parturition in cattle. *Theriogenology* 37 (4), 859–868. [https://doi.org/10.1016/0093-691x\(92\)90047-U](https://doi.org/10.1016/0093-691x(92)90047-U).
- Mattachini, G., Antler, A., Riva, E., Arbel, A., Provolo, G., 2013. Automated measurement of lying behavior for monitoring the comfort and welfare of lactating dairy cows. *Livestock Sci.* 158 (1–3), 145–150. <https://doi.org/10.1016/j.livsci.2013.10.014>.
- McGowan, J., Burke, C., Jago, J., 1999. Validation of a technology for objectively measuring behaviour in dairy cows and its application for oestrous detection. In: *Proceedings-New Zealand society of animal production*, Vol. 67, New Zealand Society of Animal Production; 1999, p. 136.
- Mee, J.F., 2004. Managing the dairy cow at calving time. *Veterinary Clinics: Food Anim. Pract.* 20 (3), 521–546.
- Ouellet, V., Vasseur, E., Heuwieser, W., Burfeind, O., Maldague, X., Charbonneau, E., 2016. Evaluation of calving indicators measured by automated monitoring devices to predict the onset of calving in holstein dairy cows. *J. Dairy Sci.* 99 (2), 1539–1548. <https://doi.org/10.3168/jds.2015-10057>.
- Ouellet, V., Vasseur, E., Heuwieser, W., Burfeind, O., Maldague, X., Charbonneau, É., 2016. Evaluation of calving indicators measured by automated monitoring devices to predict the onset of calving in holstein dairy cows. *J. Dairy Sci.* 99 (2), 1539–1548.
- Paleologo, G., Elisseeff, A., Antonini, G., 2010. Subagging for credit scoring models. *Eur. J. Oper. Res.* 201 (2), 490–499.
- Palombi, C., Paolucci, M., Stradaoli, G., Corubolo, M., Pascolo, P.B., Monaci, M., 2013. Evaluation of remote monitoring of parturition in dairy cattle as a new tool for calving management. *Bmc Veterinary Research* 9. doi: Artn 191 10.1186/1746-6148-9-191.
- Raussi, S., 2003. Human-cattle interactions in group housing. *Appl. Anim. Behav. Sci.* 80 (3), 245–262.
- Robert, B., White, B.J., Renter, D.G., Larson, R.L., 2009. Evaluation of three-dimensional accelerometers to monitor and classify behavior patterns in cattle. *Comput. Electron. Agric.* 67 (1–2), 80–84. <https://doi.org/10.1016/j.compag.2009.03.002>.
- Rutten, C., Kamphuis, C., Hogeveen, H., Huijps, K., Nielen, M., Steeneveld, W., 2017. Sensor data on cow activity, rumination, and ear temperature improve prediction of the start of calving in dairy cows. *Comput. Electron. Agric.* 132, 108–118.
- Schirmann, K., von Keyserlingk, M.A.G., Weary, D.M., Veira, D.M., Heuwieser, W., 2009. Technical note: validation of a system for monitoring rumination in dairy cows. *J. Dairy Sci.* 92 (12), 6052–6055. <https://doi.org/10.3168/jds.2009-2361>.
- Schirmann, K., Chapinal, N., Weary, D.M., Vickers, L., von Keyserlingk, M.A.G., 2013. Short communication: rumination and feeding behavior before and after calving in dairy cows. *J. Dairy Sci.* 96 (11), 7088–7092. <https://doi.org/10.3168/jds.2013-7023>.
- Schuenemann, G.M., Nieto, I., Bas, S., Galvao, K.N., Workman, J., 2011. Assessment of calving progress and reference times for obstetric intervention during dystocia in holstein dairy cows. *J. Dairy Sci.* 94 (11), 5494–5501. <https://doi.org/10.3168/jds.2011-4436>.
- Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2010. Rusboost: a hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man Cybernet. Part A-Syst. Humans* 40 (1), 185–197. <https://doi.org/10.1109/Tsmca.2009.2029559>.
- Shah, K.D., Nakao, T., Kubota, H., Maeda, T., 2007. Peripartum changes in plasma estrone sulfate and estradiol-17 beta profiles associated with and without the retention of fetal membranes in holstein-friesian cattle. *J. Reprod. Develop.* 53 (2), 279–288. <https://doi.org/10.1262/jrd.18080>.

- Streyl, D., Sauter-Louis, C., Braunert, A., Lange, D., Weber, F., Zerbe, H., 2011. Establishment of a standard operating procedure for predicting the time of calving in cattle. *J. Veterin. Sci.* 12 (2), 177–185.
- Sumi, K., Zin, T.T., Kobayashi, I., Horii, Y., 2017. A study on cow monitoring system for calving process. In: 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE). IEEE, pp. 1–2.
- The cow-calf manager: Extra attention at calving– it's so important, <https://www.sites.ext.vt.edu/newsletter-archive/livestock/aps-00_02/aps-0180.html> , accessed: 2019-08-10.
- Wright, I., White, I., Russel, A., Whyte, T., McBean, A., 1988. Prediction of calving date in beef cows by real-time ultrasonic scanning. *Vet. Rec.* 123 (9), 228–229.
- Wright, I.A., White, I.R., Russel, A.J.F., Whyte, T.K., Mcbean, A.J., 1988. Prediction of calving date in beef-cows by real-time ultrasonic-scanning. *Veterin. Rec.* 123 (9), 228–229. <https://doi.org/10.1136/vr.123.9.228>.
- Zehner, N., Niederhauser, J.J., Schick, M., Umstatter, C., 2019. Development and validation of a predictive model for calving time based on sensor measurements of ingestive behavior in dairy cows. *Comput. Electron. Agric.* 161, 62–71.