Detection and monitoring the risk level for lameness and lesions in dairy herds by alternative machine-learning algorithms

D. Warner[1,2], E. Vasseur[2], D. Lefebvre[1], R. Lacroix[1]

[1]Valacta, Dairy Production Centre of Expertise Quebec-Atlantic, Sainte-Anne-de-Bellevue, QC, Canada

[2]Department of Animal Science, McGill University, Sainte-Anne-de-Bellevue, QC, Canada

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *Machine-learning methods may play an increasing role in the development of precision agriculture tools to provide predictive insights in dairy farming operations and to routinely monitor the status of dairy cows. In the present study, we explored the use of a machine-learning approach to detect and monitor the welfare status of dairy herds in terms of lameness and lesions based on pre-recorded farm-based records. Animal-based measurements such as lameness and lesions are time-consuming, expensive and, thus, typically not collected on a routine basis. A predictive model that is suitable for routine field applications can be thus an efficient strategy to improve dairy cattle welfare. A decision tree approach was therefore used to classify the welfare status of 229 herds. Single measurements were aggregated to a composite index for lameness and lesions, scaled to percentile ranks, and expressed as low, intermediate and high risk that a herd be deficient in lameness and lesions. Routinely collected dairy herd improvement data related to milk production, milk quality, herd size, housing and reproduction were used as potential predictors of the risk level. Model accuracy based on the average of repeated 10-fold cross validation suggests that a simple decision tree algorithm was able to predict welfare level with a mean accuracy of 44%. Ensemble methods such as random forests and boosting methods slightly improved the prediction performance to some extent (up to 51% accuracy). Model specificity for herds at high welfare risk was 91% with a boosting approach, suggesting that only a small proportion of lower risk herds were misclassified as high risk herds. These results suggest that a model based on a machine-learning approach is able to detect herds with potential welfare deficiencies using routine herd data. Additional data are required to improve model performance and validate the approach. Nonetheless, a machine-learning approach may be an appropriate and powerful tool to estimate and monitor the dairy welfare status at herd level, and can be a useful decision support tool for dairy farmers.*

Keywords. *Precision monitoring, dairy herd improvement, routine herd data, animal welfare,*

machine learning.

Introduction

Dairy farming systems are gradually adopting precision dairy monitoring technologies to improve production efficiency and improve early detection of diseases. Furthermore, modern dairy farming heavily relies on the collection of a large volume and variety of data at cow and herd levels, which, often, are stored in central databases. These relatively large and complex databases typically comprise data of several thousand herds and include routinely collected farm-based data related to identification and registration, milk quality, productivity, and reproduction.

Traditional statistical methods often perform poorly with prediction and classification in large datasets (Heald et al., 2000), such as farm-based records that typically contain noisy and imbalanced information, and are often not complete. An artificial intelligence approach with machine-learning algorithms may be thus more suitable for modeling routine herd data. Machine-learning algorithms are generally capable of better handling non-linear relationships and interactions between input variables. Such an approach may also be implemented and automated in a practical setting within a central dairy record system to create alerts. Moreover, as the volume and complexity of recorded farm data is growing, especially with the use of new data loggers to monitor cow behaviour and physiology, machine-learning methods are increasingly becoming interesting to analyze routine herd data and automatically predict outcomes (Morota et al., 2018).

In the present study, a decision tree approach was used as it is particularly well suited for classification problems and is generally a useful tool for computerized decision support systems (Quinlan, 1986; Freund and Mason, 1999). With this approach, a tree-like structure with a collection of rules is generated with leaves representing the classes that individual observations are associated with. Due to their structure and easily interpretable output, decision trees are user friendly and may be used as on-farm tools to help producers make management decisions. However, a simple decision tree approach is not always considered a particularly robust classification method as predictions of a single tree are sensitive to noise in its training set. Combining decision tree with bootstrap aggregation and boosting techniques, which create multiple trees, may improve detection performance (Breiman, 1996; Freund and Shapire, 1999). Moreover, as different machine-learning methods are available, testing various algorithms is recommended to verify whether classification results can be further improved (White et al., 2018). For this reason, ensemble machine-learning methods based on bootstrap aggregation and boosting were applied to investigate whether the predictive performance of a simple decision tree can be improved.

Machine-learning algorithms have been used to detect some common diseases (Yang et al., 1999; Heald et al., 2000; Cavero et al., 2008; Sun et al., 2010), reproductive performance (Firk et al., 2003; Caraviello et al., 2006; Schefers et al., 2010; Grzesiak et al., 2010; Zaborski and Grzesiak, 2011; Shahinfar et al., 2014), calving predictions (Borchers et al., 2017), milk yield predictions (Lacroix et al., 1995; Grzesiak et al., 2003; Grzesiak et al., 2006), breeding value predictions (Shahinfar et al., 2012) in dairy cows, and even management predictions in dairy farms (Shine et al., 2018) based on a selection of pre-recorded farm-based data. Although model performance in the above-mentioned studies often suffered from the lack of insufficient input information, the approach was considered suitable for use in practice. To our knowledge, no similar approach integrating artificial intelligence and dairy cattle welfare was applied so far with the aim to detect welfare related problems on dairy farms. Because of their high prevalence on dairy farms and their long duration of cases (von Keyserlingk and Weary, 2017; Bouffard et al., 2017), the emphasis was put on detecting potential welfare deficiencies related to lameness and body lesions. Collections of these animal-based measures is particularly time-consuming and expensive (Vasseur et al., 2015), limiting their collection in routine to assess on-farm welfare. A prediction model that allows to remotely detecting cases of lameness and lesions without the need of farm visits may be therefore an efficient and promising strategy to improve dairy cattle welfare nationwide.

In the present study, on-farm animal-based measures pertaining to 229 dairy herds registered in the central dairy database were used. The dataset at hand was relatively small compared to datasets in previous studies that used machine-learning algorithms. Therefore, the objective of this exploratory study was to investigate the potential of a machine-learning approach to remotely detect the herd welfare status assessed through lameness and lesions in dairy herds using a small set of farm-based records with the aim to develop a decision support tool.

Material and Methods

Test population

The dataset used in this study is based on on-farm animal-based measures collected on 229 dairy herds in Ontario (n = 90), Quebec (n = 89) and Alberta (n = 50), Canada, from May 2011 to July 2012 using criteria and methods described by Vasseur et al. (2015). Measures included the prevalence of lameness and lesions to hock, knee and neck (Table 1). Average herd size was 112 (SD 80.0). Average standardized (305-day fat- and protein corrected) milk yield was 9713 kg (SD 855.2 kg). Of the 229 herd in the test population, 100 were housed in tie stalls and 129 in free stalls (including 110 herds that were equipped with a milking parlor and 19 with an automatic milking system). Routine herd data were extracted for the respective herds from the central dairy record of the dairy herd improvement programs (CanWest DHI in Ontario and Alberta; Valacta in Quebec) during the year before the farm visit for the animal-based measures collection. The extracted routine herd data comprised data relating to identification and registration, herd size, housing, milk production, milk quality, and reproduction.

Table 1. On-farm animal-based measures with the respective herd prevalence of the 229 test dairy herds.

Measurement	Herd prevalence (%)					
	Mean	Median	SD	Min	Max	N
Lameness	24.9	22.5	15.97	0.0	84.2	229
Hock lesions	47.8	48.7	23.21	0.0	95.0	229
Knee lesions	35.9	30.0	27.03	0.0	100.0	229
Neck lesions	18.5	7.0	23.63	0.0	100.0	229

Calculations and Statistical Analysis

Calculations and statistical analyses were conducted in R version 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria). On-farm animal-based measures were aggregated to a composite index of lameness and lesions to produce an overall assessment of the risk of deficiencies across animal-based measurements. The calculation of the composite index was based on a linear aggregation by weighting animal-based measures by their contribution to herd level welfare status (40%, 25%, 25%, and 10% for lameness and lesions on hocks, knees and neck, respectively; based on expert opinion). As the prevalence of the individual animal-based measures differed (e.g., larger prevalence for hock and knee lesions as compared to lameness and neck lesions; Table 1), the composite index would be affected more by animal-based measures with a larger prevalence. Therefore, the weighted animal-based measures used in the linear aggregation were adjusted by their median in the test population, as described in equation 1:

$$CI_{n=229} = (L \times 0.47) + (H \times 0.09) + (K \times 0.14) + (N \times 0.30) \quad (1)$$

where, *CI* = composite index of lameness and lesions, *L* = lameness, *H* = hock lesions, *K* = knee lesions, and *N* = neck lesions.

The composite index for lameness and lesions was then scaled to percentile ranks, and risk classes were created representing a low (≤ 30%), intermediate (30 > x ≤ 70) and high (> 70%) probability of a dairy herd to be potentially deficient in terms of lameness and lesions (Fig 1). The three risk classes were chosen as the main objective was to detect various levels of potential deficiencies in lameness and lesions. Furthermore, this approach is often used in practice to discriminate between high, intermediate, and low performing herds, and is a well-known concept by dairy farmers.

Collinearity was evaluated for each variable using a variance inflation factor (package mctest 1.2; Imdadullah et al., 2016), and revealed that the routine herd data were likely not strongly correlated to each other (variance inflation factor not exceeding 10; Wooldridge, 2013). Therefore, the entire dataset was used to predict dairy herd welfare level.

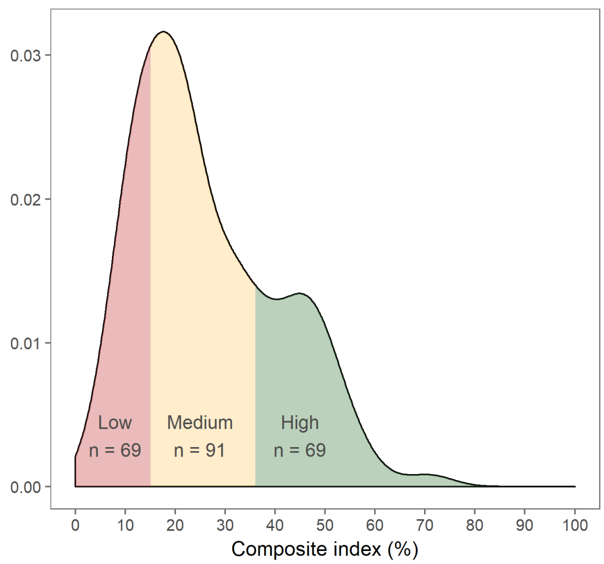

Fig 1. Distribution of the composite index of lameness and lesions with risk classes and respective class size based on 229 dairy herds (low risk: ≤ 30%; medium risk: between 30% and 70%; high risk: > 70%)

Model Development

Modeling of the welfare risk level was based on both a simple classification tree and an ensemble method, which constructs and combines multiple trees to yield a single consensus prediction. For a simple classification tree, the classification and regression tree (CART) algorithm by Breiman et al. (1984) based on recursive portioning was used. With this approach a tree-like structure is generated with a collection of rules that create branches, where the ends are terminal nodes at which point observations are classified into one of the three risk classes. Splitting stops when the algorithm detects that no further gain can be made.

Next to a simple decision tree, ensemble methods based on bootstrap aggregation and boosting were used to investigate whether the predictive performance could be improved compared to simple decision trees. Bootstrap aggregation repeatedly selects a random sample with replacement of the training set and fits multiple trees to these samples (i.e., each tree is shown a different training set). The random forest algorithm by Breiman (2001), which implements this method but further randomly selects variables to prevent trees from becoming correlated, was used in the present study. Next to it, boosting methods were used as they generally improve model accuracy and can be used in conjunction with many other types of machine-learning algorithms. This approach is essentially based on creating several simplified trees, each using merely a fraction of all available variables. Each successive tree adapts to those classification instances where previous trees failed by using a different set of variables. In the present study, the adaptive boosting (AdaBoost) algorithm by Freund and Schapire (1999) was applied as it is a popular boosting algorithm that can adapt to previous trees. However, this method is sensitive to noisy data and outliers (Freund and Mason, 1999; Mason et al., 1999), which are typically found in routine herd data. Therefore, another popular boosting algorithm, the gradient boosting machine (GBM) algorithm by Friedman (2001) was applied, which may better handle noisy data and outliers. Lastly, the extreme gradient boosting (XGBoost) algorithm by Chen and Guestrin (2016) was applied as this method was specifically developed to further optimize computational resources for gradient boosted tree algorithms and might be, thus, useful for a potential implementation in a practical setting within a central dairy record system.

Models were developed in R version 3.4.4 (R Foundation for Statistical Computing, Vienna, Austria), using the modelling wrapper package caret 6.0-79 (Kuhn, 2008). Missing data were imputed by predictive mean matching based on the approach described by van Buuren (2012). This method constructs matching cases with missing data to similar cases with data present, and generally handles well variables that are not normally distributed. Data imputation was done using the package mice 2.46.0 (van Buuren and Grootshuis-Oudshoorn, 2011). Extreme values in the herd data were removed after visual inspection. Models were prepared and the respective tuning parameters were adjusted using 10-fold cross-validation with three iterations. For each iteration, a model was prepared using nine splits of the data set, and the model was evaluated on the remaining part of the data set (i.e., one split). Model accuracy was estimated based on the average of the 10-fold cross validation repeated 3 times, and the accuracy distribution (based on 30 values) was evaluated. This approach was found useful to analyze results for small data sets (Pietersma et al., 2003) as all observations are used for both developing and validating the model, and each observation is used for model validation exactly once.

As one potential application of such a predictive model is the detection and identification of high and low performing herds in terms of lameness and lesions, model performance was further evaluated by estimating the proportion of correctly identified or correctly rejected herds for each of the three risk classes (high, medium, low). For this reason, data were randomly partitioned into a train data subset (70% of all data) to prepare models and a test data subset (30%). The risk level classes were partitioned similarly among the train and test data subsets (Table 2). However, both data sets had a slightly larger number of observations for the medium risk class as compared the low and high risk class. Average herd size and milk yield were similar between the train and the test data subsets. Likewise, housing systems were similarly partitioned between the subsets suggesting that observations were randomly split between the train and test data subsets. Model sensitivity and specificity were calculated. In the present study based on a multinomial classification, sensitivity represents the percentage of herds correctly identified within the respective risk class (e.g. a high risk herd identified as such; equation 2), and specificity represents the percentage of herds correctly rejected as not being within the respective risk class (e.g. a herd not at high risk identified as such; equation 3:

$$Sensitivity = \frac{TP}{TP+FN} \tag{2}$$

$$Specificity = \frac{TN}{TN+FP} \tag{3}$$

where, TP = true positive (i.e., correctly identified case), FP = false positive (i.e., incorrectly identified case), TN = true negative (i.e., correctly rejected case), FN = false negative (i.e., incorrectly rejected case).

Table 2. Number of observations and herd characteristics in the total, training and test data subset.

Item	Total data set	Subsets	
		Train	Test
Observations	229	162	67
Observation per risk class[1]			
Low	69	49	20
Medium	91	64	27
High	69	49	20
Housing system			
Tie stalls	100	70	30
Free stalls	129	92	37
Average herd size	112	111	114
Average standardized milk yield[2]	9713	9715	9700

[1]Risk for lameness and lesions on herd level
[2]305-day fat- and protein-corrected milk

Results and Discussion

Multiple machine-learning methods based on a decision tree approach were tested to classify the welfare status of dairy herds with the focus on lameness and lesions using farm-based records related to milk production, milk quality, stock number, housing and reproduction. Model accuracy based on the average of the repeated 10-fold cross validation suggests that a simple decision tree algorithm (CART) was able to predict welfare level with a mean accuracy of 44% (Fig 2). Ensemble methods seem to improve the prediction performance to some extent. A random forest algorithm had a mean accuracy of 49%. Boosting seemed to slightly further improve prediction performance, in particular using a gradient boosting approach with a mean model accuracy of 51% for a GBM algorithm. Next to model accuracy, the accuracy distribution should be taken into account to evaluate predictive performance of a model. For instance, Fig 2 suggests that maximum model accuracy may be as high as 74% for a GBM model. However, the prediction outcome is far less variable for a CART and an AdaBoost algorithm. Despite a lower model accuracy, an approach by CART or AdaBoost might be thus still of interest in practice as their predictive behavior with a future unseen dataset can more reliably estimated.

The results shown in Fig 2 suggest that the predictive performance of the machine-learning algorithms used in the present study was comparable. A pairwise t-test with Bonferroni *P*-value adjustment confirmed that a GBM algorithm had a significantly ($P = 0.007$) higher model accuracy than a CART algorithm, but the model accuracy among all other algorithms was not significantly different ($P > 0.08$). These results suggest that a CART algorithm can be still a powerful approach. Vázquez Diosdado et al. (2015) observed that accuracy of the decision tree algorithm matched the performance of more computationally intensive algorithms (i.e., hidden Markov models and support vector machines) in classifying behavioral activities in dairy cows based on specific information from neck-mounted accelerometers.

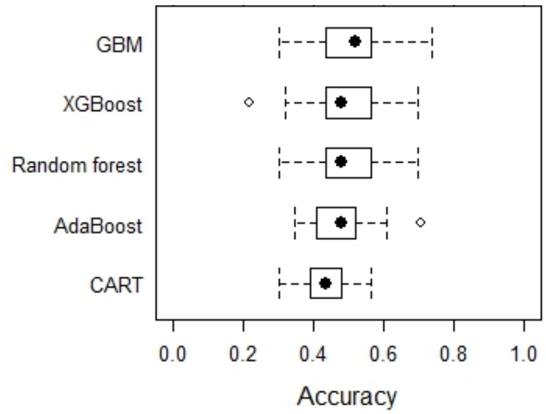

Fig 2. Model accuracy for the tested machine-learning algorithms (GBM: gradient boosting machine; XGBoost: extreme gradient boosting; AdaBoost: adaptive boosting; CART: classification and regression tree). Boxplots display the respective interquartile range (box), median (solid bullet), minima and maxima (whiskers), and extreme values (empty bullet) for accuracy.

It is expected that the prediction performance can be further improved by including a larger variety of farm-based records and more informative parameters. For instance, de Vries et al. (2014) reported a model accuracy of 72% to 81% for lameness and lesions based on multivariate regression but using parameters that were not available in the present study (e.g. heifer udder infections, days in milk, grazing management, on-farm mortality for cows after 60 days in milk, somatic cell count for cows at various stages of lactation, average service per cow). In the present study, the most important variables selected by the algorithms were percentage of culled cows, stock number and body weight of lactating cows that appeared in all 5 models, proportion of primiparous cows (appeared in 4 models), and milk urea nitrogen content (appeared in 3 models) (Table 3). Collection of more specific indicators known to be associated to lameness and lesions may further improve model performance. For instance, lying behavior (Ito et al., 2010) and body condition score (Solano et al., 2015) were found to be good indicators of lameness in dairy cows. However, collection of such animal-based measures are currently not done on routine basis. Nonetheless, national initiatives on herd welfare status evaluation (e.g., animal care assessment through proAction currently on-going by Dairy Farmers of Canada; Dairy Farmers of Canada, 2015), and future prospects of animal-based estimates with digital images or activity sensors (e.g., body condition scores estimations; Bewley et al., 2008) suggest that there is ample opportunities for improving model performance through the collection of more specific animal-based measures.

Table 3. Variable selection (10 most important variables) used to fit the respective model for welfare risk predictions.

Algorithm[1]	Variables
CART	Body weight of lactating cows, stock number, proportion of primiparous cows, housing system, percentage of culled cows, milk urea nitrogen content, average calving interval, proportion of lactating cows
Random forest	Stock number, body weight of lactating cows, temperament score, proportion of primiparous cows, percentage of culled cows, milk urea nitrogen content, percentage of culled cows until 30 days in milk, percentage of sold cows, Transition Cow Index, turnover rate for heifers
AdaBoost	Stock number, housing system, body weight of lactating cows, proportion of primiparous cows, percentage of culled cows for involuntary reasons, percentage of culled cows, percentage of stillborn calves
GBM	Stock number, body weight of lactating cows, temperament score, mortality rate, percentage of culled cows, turnover rate, proportion of primiparous cows, milk urea nitrogen content, days until first breeding, percentage of stillborn calves
XGBoost	Stock number, percentage of culled cows until 30 days in milk, percentage of primiparous cows, body weight of lactating cows, temperament score, percentage of culled cows, average calving interval, milk management index, days until first breeding, inbreeding coefficient

[1]CART: classification and regression tree; AdaBoost: adaptive boosting; GBM: gradient boosting machine; XGBoost: extreme boosting

Model performance was further evaluated in terms of sensitivity (i.e., correctly identified herds) and specificity (i.e., correctly rejected herds) for each of the three welfare risk classes. As shown in Table 2, the proportion of cases available for each class was slightly unbalanced with the medium risk class as the majority class. In general, machine-learning algorithms do not cope well with unbalanced classifications, and over-sampling and under-sampling techniques exist to achieve better classifier performance, but the imbalance was considered acceptable for use with machine-learning algorithms in the present study. Model sensitivity and specificity varied somewhat among the various models (Table 4). In the present study, specificity may be more relevant in view of a potential application in practice as dairy herds at high welfare risk need to be solicited and examined more closely. That is, rather than correctly identifying a large amount of herds of low or high risk of lameness and lesions (i.e., through a high sensitivity), it may be more relevant to avoid that herds are incorrectly classified into a low or high risk class (i.e., through a high specificity) and, thus, prevent that these herds are wrongly examined in detail at a later stage. In line with results in Fig 2, a GBM model seemed to be more specific within the low and high risk classes as compared to a simple decision tree. As there is always a trade-off between specificity and sensitivity, model sensitivity within the low and high risk classes decreased for a GBM model. The other models do not appear to be particularly more specific within the low and high risk class than a simple decision tree algorithm, in line with results in Fig 2. It should be noted that the results in Table 4 cannot be directly compared to those in Fig 2. The latter is based on a three-time repeated 10-fold cross-validation; thus, the entire dataset was used with only one fold withhold for validation and model accuracy was averaged across 30 values. On the other hand, sensitivity and specificity are calculated based on a smaller dataset (70%) and validated on the remaining dataset to obtain a better insight on how each model performs within each of the three risk classes.

Table 4. Performance of machine-learning algorithms across the respective lameness and lesion risk classes.

		Risk class		
Algorithm[1]	Model performance	Low	Medium	High
CART	Sensitivity	0.60	0.44	0.75
	Specificity	0.70	0.83	0.85
GBM	Sensitivity	0.50	0.59	0.50
	Specificity	0.79	0.58	0.91
AdaBoost	Sensitivity	0.20	0.48	0.50
	Specificity	0.79	0.50	0.79
XGBoost	Sensitivity	0.45	0.59	0.55
	Specificity	0.79	0.68	0.83
Random forest	Sensitivity	0.45	0.48	0.50
	Specificity	0.74	0.58	0.87

[1]CART: classification and regression tree; AdaBoost: adaptive boosting; GBM: gradient boosting machine; XGBoost: extreme boosting

Conclusion

This paper presented results on the use of a machine-learning approach to detect dairy herd welfare level based on lameness and lesions. The results suggest that in principle, machine-learning algorithms can be used to classify welfare level of dairy herds but further efforts need to be done to improve model performance. Model accuracy of a simple decision tree (44%) was slightly improved with ensemble methods based on bootstrap aggregation and boosting (up to 51% accuracy). A comparison of a machine-learning approach to multivariate regression might be interesting to assess whether traditional statistical analysis performs similarly with data derived from routinely collected farm-based records. Model performance suffered from the relatively small set of observations available and from the lack of a more focused set of farm-based records in the present study. Additional information more closely related to lameness and lesions may improve the performance of the model. A follow-up study with a larger amount of observations for lameness and lesions, and more relevant set of routinely collected data is needed to determine whether dairy herd welfare status can be more accurately identified using a machine-learning approach. In addition, further studies are required to investigate whether machine-learning methods can be also used to predict the welfare status of individual animals rather than on herd level. Nonetheless, the approach proposed in the present study offers opportunities to integrate artificial intelligence and dairy cattle welfare within a computerized information system. Such an approach based on classification of farm-based records into multiple welfare categories is suitable for routine field applications to create automatic alerts for herds with welfare deficiencies, and can be useful as a decision support system for dairy farmers.

Acknowledgements

This study was funded by a project from Agriculture & Agri-Food Canada (Ottawa, ON, Canada) and the Ministry of Agriculture, Fisheries and Food (Quebec, QC, Canada) under the framework “Growing Forward 2” (grant number 17-3-07), as well as funding from Natural Sciences and Engineering Research Council of Canada, Novalait, Dairy Farmers of Canada, and Valacta through Vasseur’s Industrial Research Chair. Data used in this paper were collected as part of a previous study funded by Agriculture & Agri-Food Canada and Dairy Farmers of Canada (Ottawa, ON, Canada) as part of the Dairy Science Cluster initiative, as well as Fonds de recherche nature et technologies (Quebec, QC, Canada), Ministere de l'Agriculture, des Pecheries et de l'alimentation du Quebec (Quebec, QC, Canada), Novalait Inc. (Quebec, QC, Canada), and Agriculture & Agri-Food Canada (Ottawa, ON, Canada) as part of the partnership program for innovation in dairy production and transformation. We thank the leads of this previous study in each province where data collection was conducted: Doris Pellerin (Université Laval, QC), Derek Haley (University of Guelph, ON), and Karin Orsel (University of Calgary, AB).

References

Bewley, J. M., Peacock, A. M., Lewis, O., Boyce, R. E., Roberts, D. J., Coffey, M. P., Kenyon, S. J., & Schutz, M. M. (2008). Potential for estimation of body condition scores in dairy cattle from digital images. *Journal of Dairy Science*, 91(9), 3439–3453.

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. *Journal of Dairy Science*, 100, 1–11.

Bouffard, V., de Passillé, A. M., Rushen, J., Vasseur, E., Nash, C., Haley, D., & Pellerin, D. (2017). Impact of following recommendations for tie stall configuration on neck and leg lesions, lameness, cleanliness and lying time in dairy cows. *Journal of Dairy Science*, 100, 2935-2943.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 132–140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Monterey, CA: Wadsworth & Brooks.

Caraviello, D. Z., Weigel, K. A., Craven, M., Gianola, D., Cook, N. B., Nordlund, K. V., Fricke, P. M., & Wiltbank, M. C. (2006). Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. *Journal of Dairy Science*, 89, 4703–4722.

Cavero, D., Tölle, K. H., Henze, C., Buxadé, C., & Krieter, J. (2008). Mastitis detection in dairy cows by application of neural networks. *Livestock Science*, 114, 280–286.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY: ACM.

Dairy Farmers of Canada (2015). proAction® Implementation. http://www.proaction.quebec/wp-content/uploads/2016/05/proActionImplementationENG.pdf. Accessed 27 March 2018.

de Vries, M., Bokkers, E. A. M., van Schaik, G., Engel, B., Dijkstra, T., & de Boer, I. J. M. (2014). Exploring the value of routinely collected herd data for estimating dairy cattle welfare. *Journal of Dairy Science*, 97, 715–730.

Firk, R., Stamer, E., Junge, W., Krieter, J. (2003). Improving oestrus detection by combination of activity measurements with information about previous oestrus cases. *Livestock Production Science*, 82, 97–103.

Freund, Y., & Mason, L. (1999). The alternating decision tree learning algorithm. In I. Bratko, & S. Dzeroski (Eds.), Proceedings of the 16th International Conference on Machine Learning (pp. 124–133). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Freund, Y., & Schapire. R., E. (1999). A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5), 771–780.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.

Grzesiak, W., Błaszczyk, P., & Lacroix, R. (2006). Methods of predicting milk yield in dairy cows – Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and Electronics in Agriculture*, 54, 69–83.

Grzesiak, W., Lacroix, R., Wójcik, J., & Błaszczyk, P. (2003). A comparison of neural network and multiple regression predictions for 305-day lactation yield using partial lactation records. *Canadian Journal of Animal Science*, 83, 307–310.

Grzesiak, W., Zaborski, D., Sablik, P., Żukiewicz, A., Dybus, A., & Szatkowska, I. (2010). Detection of cows with insemination problems using selected classification models. *Computers and Electronics in Agriculture*, 74: 265-273.

Heald, C. W., Kim, T., Sischo, W. M., Cooper, J. B., & Wolfgang, D. R. (2000). A computerized mastitis decision aid using farm-based records: an artificial neural network approach. *Journal of Dairy Science*, 83, 711–720.

Imdadullah, M., Aslam, M., & Altaf, S. (2016). mctest: an R package for detection of collinearity among regressors. *The R Journal*, 8(2), 499–509.

Ito, K., von Keyserlingk, M. A. G., LeBlanc, S. J., & Weary, D. M. (2010). Lying behavior as an indicator of lameness in dairy cows. *Journal of Dairy Science*, 93(8), 3553–3560.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26.

Lacroix, R., Wade, K. M., Kok, R., & Hayes, J. F. (1995). Prediction of cow performance with a connectionist model. *Transactions of the ASAE*, 38(5), 1573–1579.

Mason, L., Baxter, J., Bartlett, P. L., & Freen, M. (1999). Boosting algorithms as gradient descent in function space. In S. A. Solla, T. K. Leen, & K. Müller (Eds.), *Advances in Neural Information Processing Systems* (pp. 512–518). Cambridge, MA: MIT Press.

Morota, G., Ventura, R. V., Silva, F. F., Koyama, M., & Fernando, S. C. (2018). Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *Journal of Animal Science*, 96, 1540–1550.

Pietersma, D., Lacroix, R., Lefebvre, D., & Wade, K. M. (2003). Performance analysis for machine-learning experiments using small data sets. *Computers and Electronics in Agriculture*, 38, 1–17.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1, 81–106.

Schefers, J. M., Weigel, K. A., Rawson, C. L., Zwald, N. R., & Cook, N. B. (2010). Management practices associated with conception rate and service rate of lactating Holstein cows in large, commercial dairy herds. *Journal of Dairy Science*, 93, 1459–1467.

Shahinfar, S., Mehrabani-Yeganeh, H., Lucas, C., Kalhor, A., Kazemian, M, & Weigel, K. A. (2012). Prediction of breeding values for dairy cattle using artificial neural networks and neuro-fuzzy systems. Computational and Mathematical Methods in Medicince, 2012, 127130, doi: 10.1155/2012/127130.

Shahinfar, S., Page, D., Guenther, J., Cabrera, V., Fricke, P., & Weigel, K. (2014). Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. *Journal of Dairy Science*, 97, 731–742.

Shine, P., Murphy, M. D., Upton, J., & Scully, T. (2018). Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms. *Computers and Electronics in Agriculture*, 150, 74–87.

Solano, L., Barkema, H. W., Pajor, E. A., Mason, S., LeBlanc, S. J., Zaffino Heyerhoff, J. C., Nash, C. G. R., Haley, D. B., Vasseur, E., Pellerin, D., Rushen, J., de Passillé, A. M., & Orsel, K. (2015). Prevalence of lameness and associated risk factors in Canadian Holstein-Friesian cows housed in freestall barns. *Journal of Dairy Science*, 98, 6978–6991.

Sun, Z., Samarasinghe, S., & Jago, J. (2010). Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks. *Journal of Dairy Research*, 77, 168–175.

van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Boca Raton, FL: CRC Chapman & Hall.

van Buuren, S, & Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1-67.

Vasseur, E., Rushen, J., Gibbons, J., Pellerin, D., Pajor, E., Lefebvre, D., & de Passillé, A. M. (2015). An advisory tool to help producers improve cow comfort. *Journal of Dairy Science*, 98, 1-11.

Vázquez Diosdado, J. A., Barker, Z. E., Hodges, H. R., Amory, J. R., Croft, D. P., Bell, N. J., & Codling E. A. (2015). Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system. *Animal Biotelemetry*, 3(1), doi: 10.1186/s40317-015-0045-8.

von Keyserlingk, M. A. G., & Weary D. M. (2017). A 100-year review: animal welfare in the Journal of Dairy Science—The first 100 years. *Journal of Dairy Science*, 100, 10432–10444.

White, B. J., Amrine, D. E., & Larson, R. L. (2018). Data to decisions. *Journal of Animal Science*, 96, 1531–1539.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach.* Fifth Edition. Mason, OH: Southwestern (p. 98).

Yang, X. Z., Lacroix, R., & Wade, K. M. (1999). Neural detection of mastitis from Dairy Herd Improvement records. *Transactions of the ASAE*, 42(4), 1063–1072.

Zaborski, D., & Grzesiak, W. (2011). Detection of difficult calvings in dairy cows using neural classifiers. *Archiv Tierzucht*, 54(5), 477-489.