DOI:10.3220/LBF1515508151000

Use of random forest for dystocia detection in dairy cattle

Daniel Zaborski*, Witold S. Proskura*, Wilhelm Grzesiak*, Iwona Szatkowska*, and Magdalena Jędrzejczak-Silicka*

Abstract

The aim of the present study was to illustrate the predictive performance of random forest (RF) used for dystocia detection in dairy cattle. A total of 1,342 and 1,699 calving records of Polish Holstein-Friesian Black-and-White heifers and cows were used. Five or ten predictor variables were included in the RF models for heifers and cows, respectively. The output variable was calving class. The proportion of correctly detected easy, moderate and difficult calving events in heifers on the independent test set was 39.64 %, 57.39 % and 83.64 %, respectively. The total accuracy was recorded as 60.12 %. The corresponding values for cows were 69.39 %, 67.61 %, 0 % and 66.04 %. The most significant predictors for heifers were sire's rank and calving age, whereas those for cows additionally included: daily milk yield for the preceding lactation and the length of calving interval. The RF model developed in the present study was characterized by a high percentage of correctly diagnosed difficult calving events in heifers. However, it was completely unable to correctly detect dystocia in cows. The use of more influential predictor variables for cows in future research is especially important.

Keywords: *data mining, calving, dairy cow, dairy heifer, detection, gains chart*

Zusammenfassung

Nutzung von Random Forest zur Detektion der Dystokie bei Milchvieh

Ziel der Arbeit war, die Vorhersageleistung des Random Forest Klassifikationsverfahrens (RF) zur Erkennung der Dystokie beim Milchvieh darzustellen. Zu diesem Zweck wurden entsprechend 1342 und 1699 Aufzeichnungen über das Abkalben von Färsen und Kühen der polnischen, schwarzweißen Holstein-Friesian Rasse ausgewertet. In den RF-Modellen für Färsen und Kühe wurden entsprechend fünf bzw. zehn Prädiktoren berücksichtigt. Die Ausgangsvariable war die Klasse des Abkalbens. Das Verhältnis der erkannten leichten, moderaten und komplizierten Kalbungen von Färsen betrug beim unabhängigen Test-Set entsprechend 39,64 %, 57,39 % und 83,64 %. Die allgemeine Treffgenauigkeit erzielte den Wert von 60,12 %. Die entsprechenden Werte für Kühe waren 69,39 %, 67,61 %, 0 % und 66,04 %. Die am meisten signifikanten Prädiktoren für das Abkalben der Färsen waren: die Klasse der Bullen und das Alter des Abkalbens. Dagegen umfassten die wichtigsten Prädiktoren beim Abkalben von Kühen zusätzlich die täglichen Milchertrag in der vorherigen Laktation und das Intervall zwischen den Kalbungen. Das eingesetzte RF-Modell zeichnete sich durch einen hohen Prozentsatz korrekt diagnostizierter komplizierter Kälbungen bei Färsen aus. Jedoch war dieses Modell völlig ungeeignet, Dystokie bei Kühen zu erkennen. Die Nutzung einflussreicherer Prädiktoren für Kühe ist besonders wichtig für zukünftige Forschung.

Schlüsselworte: *Data-Mining, Abkalben, Milchkuh, Milchfärse, Detektion, Gewinndiagramm*

* West Pomeranian University of Technology, Szczecin, Department of Ruminants Science, Klemensa Janickiego 29, 71-270 Szczecin, Poland

Contact: daniel.zaborski@zut.edu.pl

1 Introduction

Dystocia (difficult calving) in cattle results in many adverse consequences (Mee et al., 2011; Rutten et al., 2017). These include: the loss of animals, reduced productive and reproductive performance and compromised animal welfare, all of which generate increased labor and veterinary costs contributing to lower farm incomes (Kumar et al., 2017; Martin-Collado et al., 2017). Moreover, approx. half of severe dystocic parturitions result in stillbirths and calves born from difficult calving are 1.5 times more likely to develop different diseases during the first 120 days of life (Kovács et al., 2017; Rutten et al., 2017). The recent estimates of dystocia prevalence in various counties (Miedema et al., 2011) reflect that it typically ranges between 2 and 7 %; however, it is higher in the USA (around 13 %). The risk of dystocia is influenced by two main components: direct (associated with the traits of the calf such as its body size, hormonal balance, body weight) and maternal (associated with the traits of the dam such as pelvic measurements, expression of maternal behavior, responsiveness to parturition signaling etc.) (Martin-Collado et al., 2017; Rutten et al., 2017). Taking into account the negative sequels of dystocia, the prediction of its occurrence is very beneficial in animal production practice. Such a prediction is possible with the use of more sophisticated statistical models, especially those from the field of data mining.

One of the powerful data mining classification methods, developed on the basis of decision trees, is a random forest (RF) (Breiman, 2001), especially suited for the datasets containing many predictors. It is often considered as an improvement over the classification and regression trees (CART) algorithm, which, in its basic form, uses only a single tree to generate the final prediction of an output (response) variable (Naidoo et al., 2012). In the case of RF, many decision trees are constructed using a random selection of training records and a subset of predictor variables at each iteration. The final prediction of the values of the output variable is based on voting, performed by all the component trees in the model. Thus, prediction accuracy can be increased (Grossmann et al., 2010). There are two main sources of randomization in the RF growing procedure: the first one results from the bootstrap sampling of the training data set and the second one arises from the random subsampling of the predictor variables at each node (Brieuc et al., 2015). Moreover, in order to additionally decrease the probability of model overfitting, the whole data set is randomly divided into a training and validation set, both of which are employed to estimate a classification error. The addition of new trees to the model stops when the error on the validation set starts rising (Nisbet et al., 2009). The recent applications of RF in cattle farming include: the prediction of calving events in Holsteins from automatically recorded data (activity, lying, and rumination) (Borchers et al., 2017), the prediction of bull behavior on multiple-sire pastures from accelerometer data (Abell et al., 2017), the prediction of behavioral states (walking, grazing, and resting) in Swiss dairy cattle from the global positioning system data (Homburger et al., 2014) as well as health status (mastitis, reproductive, and metabolic disorders) in dairy cows (Gaddis et al., 2016). In addition, RF was used for the identification of bovine viral diarrhea virus (Machado et al., 2015), *Fasciola hepatica* infections in European dairy herds (Selemetas et al., 2015), detection of estrus in Holstein cows based on sensor-recorded data (Dolecheck et al., 2015), and bovine spongiform encephalopathy (Menze et al., 2007). Finally, RF was utilized for the classification of bull semen samples from several chemically determined elements (Aguiar et al., 2012).

Therefore, the aim of the present study was to illustrate the predictive performance of RF used for dystocia detection in dairy cattle.

2 Materials and methods

A total of 1,342 and 1,699 calving records of Polish Holstein-Friesian Black-and-White heifers and cows were used. The data were collected between the years 2002 and 2013 from four commercial dairy farms located in the West Pomeranian Province. The animals were kept in free-stall barns and fed a total mixed ration. An initial set of 1,656 and 2,136 calving records (for heifers and cows, respectively) was reduced after editing for missing and erroneous values and outliers. The following predictors were included in the RF model for heifers:

X_1 – AGE – calving age (in months),
X_2 – SIRE – the rank of the heifer's sire based on the mean calving difficulty scores of its daughters (in scores),
X_3 – FARM – the category of the farm where the heifer was kept based on its average milk production (lower: <10,200 kg milk or higher: ≥10,200 kg milk),
X_4 – SEASON – calving season (autumn-winter from October to March and spring-summer from April to September),
X_5 – SEX – calf sex (male, female). The RF model for cows additionally included:
X_6 – DMY – the mean daily milk yield for the preceding lactation (in kg),
X_7 – CIN – preceding calving interval (in days),
X_8 – LACT – lactation number,
X_9 – PDIF – preceding calving difficulty (easy, moderate or difficult),
X_{10} – MAST – udder diseases during pregnancy.

The sire's rank (SIRE) was calculated after the identification of each sire's daughters from each of the four farms. The original calving difficulty scores (on the scale from 1 to 5) of the sire's daughters were then averaged; and the sires were ranked according to an increasing average calving difficulty score (with the lowest rank indicating the sire with the easiest calving events, and the highest rank indicating the sire with the most difficult calving events). Farm category (FARM) was determined using the k-means clustering method. Two farms were assigned to the lower category and two farms to the higher one. The response (output) variable Y was calving difficulty class (easy, moderate and difficult). In order to allow comparisons with previous studies that utilized only two categories of calving difficulty, an alternative scoring including two classes (easy/moderate vs. difficult) was also used in the present work. A more detailed description of the dataset is presented elsewhere (Zaborski et al., 2017, 2016).

Before the stage of model construction, the whole data set was randomly partitioned into a training set (L; 1,006 and 1,275 records for heifers and cows, respectively) for model construction and a testing set (T; 336 and 424 records for heifers and cows, respectively) for the objective verification of its predictive performance. In the construction of the RF model, equal costs of misclassification and the a priori probabilities estimated from the training sample were employed. The number of predictor variables randomly selected for individual splits in the component trees of the RF model was three for heifers and four for cows. The proportion of training cases (calving events) used for growing single trees was 50 % both for heifers and cows. The component trees were added to the model until reaching the lowest fraction of misclassification on the validation set (30 % calving records separated from the training set and used to prevent overfitting).

In order to better depict the predictive performance of the RF models, cumulative gains charts were also plotted. These plots called sometimes "banana charts" display the relationship between cumulative gains (the proportion of calving records from a distinguished category among all the records belonging to this category) and the percentage of records predicted by the model to belong to this category in the whole test set (Burez and Van den Poel, 2009). The diagonal line crossing the (0,0) and (1,1) points (the baseline) reflects a "random" model or, in other words, no model at all, whereas the curves located further above from this baseline are preferred [the closer the curve to the (0,1) point, the better the model: the banana is then "fatter"] (Ha et al., 2005).

Finally, the importance analysis was conducted to show the most influential predictor variables. As a result of the RF model building, the importance scores for individual predictor variables were obtained. These scores may be interpreted in a similar way as the p-values from statistical tests (although in an opposite direction – the higher the score, the more important a given variable) and used for the ordering of predictor variables according to their relative importance to the RF model (Yao et al., 2013). All the statistical computations were performed using Statistica software (v. 13, Dell Inc., Tulsa, OK, USA).

3 Results

3.1 Model quality

The final RF model for heifers consisted of 680 component trees. The percentage of correctly diagnosed easy, moderate and difficult calving events on the L set was 45.73 %, 62.60 %, 77.60 %, respectively. The total accuracy (the proportion of correctly indicated calving events from all classes) was 61.83 %. In the case of only two classes of calving difficulty, the proportion of correctly classified easy and difficult calving events in heifers was 92.31 % and 65.93 %, respectively, and the overall accuracy was 84.00 %. The final RF model for cows comprised 520 individual trees. The proportion of correctly classified easy, moderate and difficult calving events on the L set was 68.09 %, 72.88 % and 0 %, respectively. The total accuracy was 68.31 %. In the case of only two classes of calving difficulty, the proportion of correctly classified easy and difficult calving events in cows was 100 % and 0 %, respectively. Consequently, the overall accuracy for the two-class system was 96.63 %.

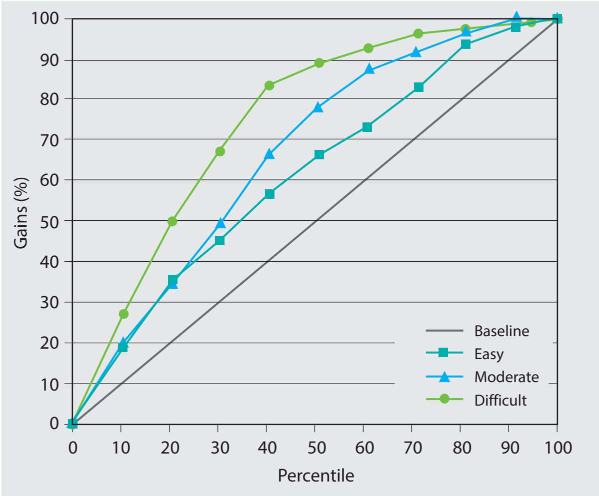

Figure 1
The cumulative gains chart for calving difficulty in heifers

3.2 Predictive performance

At the next stage of the present study, both models (for heifers and cows) were verified on the independent T set. The proportion of correctly detected easy, moderate and difficult calving events in heifers was 39.64 %, 57.39 % and 83.64 %, respectively. The accuracy was 60.12 %. In the case of only two classes of calving difficulty, the proportion of correctly identified easy and difficult calving events in heifers was 86.28 % and 70.91 %, respectively. The overall accuracy in this case was 81.25 %. The percentage of correctly detected easy, moderate and difficult calving events in cows was 69.39 %, 67.61 %, and 0 %, respectively, with a total accuracy of

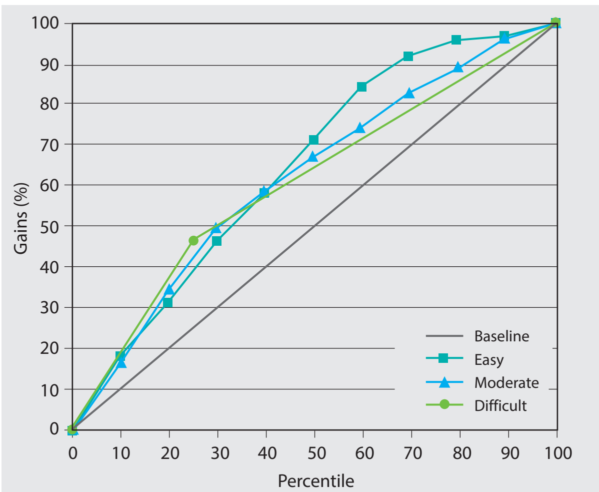

Figure 2
The cumulative gains chart for calving difficulty in cows

66.04 %. In the case of only two classes of calving difficulty, the proportion of correctly identified easy and difficult calving events was 100 % and 0 %, respectively. The overall accuracy for the two-class system was 96.46 %. The cumulative gains charts for heifers and cows (for the three categories of calving difficulty) are presented in Figure 1 and 2, respectively.

3.3 Predictor importance

The importance of individual predictors for heifers and cows is shown in Figure 3 and 4, respectively. The most influential predictor variables for heifers were SIRE and AGE and those for cows included: SIRE, AGE, DMY and CIN. The rest of them were of much lower importance.

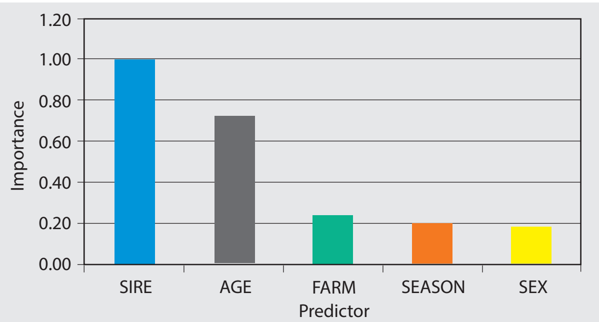

Figure 3
Predictor importance for the random forest model in heifers

4 Discussion

The quality of the RF model for heifers obtained in the present study was moderate (accuracy of approx. 62 %). However, the fact that it predicted dystocia cases most precisely (approx. 78 % correctly classified cases) is especially noteworthy. Almost the same accuracy was reported by Piwczyński et al. (2013) who applied CART and CHAID to calving difficulty classification (with four categories) in Polish Holstein-Friesian Black-and-White cows. Comparable outcomes (accuracy ranging from 50 to 60 %) were also reported by Johnson et al. (1988) who used discriminant function analysis for calving difficulty prediction (five classes) in Hereford heifers. The final value of this probability in the cited study depended on the set of predictor variables (prebreeding or precalving) included in the model and the number of calving difficulty classes. It amounted to approximately 86 % in the case in which only two classes of calving (normal vs. difficult) were considered. This value was obviously much higher compared to the value obtained in the present work for the model with three classes. But the RF model containing the binary response variable (easy/moderate vs. difficult) used in the present study to allow comparisons with other studies had almost the same accuracy (84 %). Also, Arthur et al. (2000) reported higher accuracy (85 to 92 %) in their study on dystocia detection in Angus heifers by means of discriminant function analysis. However, this result was achieved for only two classes of calving difficulty (easy or difficult). When compared to the model with two calving categories in our study (accuracy equal to 84 %), this outcome was very similar to ours. The quality of the model for cows (accuracy on the L set equal to approx. 68 %) in the present work was slightly better than that for heifers but, in general, it was also moderate. Like for heifers, its value was similar to that (62 %) reported by Piwczyński et al. (2013). A very negative result recorded in our study was the total inability of the RF model to appropriately diagnose difficult calving events in cows. It could have been caused by the very low frequency of such records in the data set (approx. 3 %).

At the next stage of the present work, the predictive performance of the RF models for heifers and cows was confirmed on the independent T set, whose calving records had not been used previously during the model development phase. The moderate abilities of the constructed models to correctly indicate calving events from three groups observed on the L set were, in general, confirmed on the T set. And so, the accuracy of RF for heifer calving events (approx. 60 %) was markedly lower than the values (73 to 90 %) reported by Arthur et al. (2000). However, the ability of the discriminant function model described in the cited study to correctly indicate difficult calving events (expressed as sensitivity) was relatively low (0 to 40 %), whereas in the present work, the proportion of properly detected dystocic records in heifers on the T set was much higher (about 84 %), which is especially noteworthy. On the contrary, the percentage of correctly classified easy calving events (specificity) in the study by Arthur et al. (2000) ranged from approx. 80 to 100 %, while

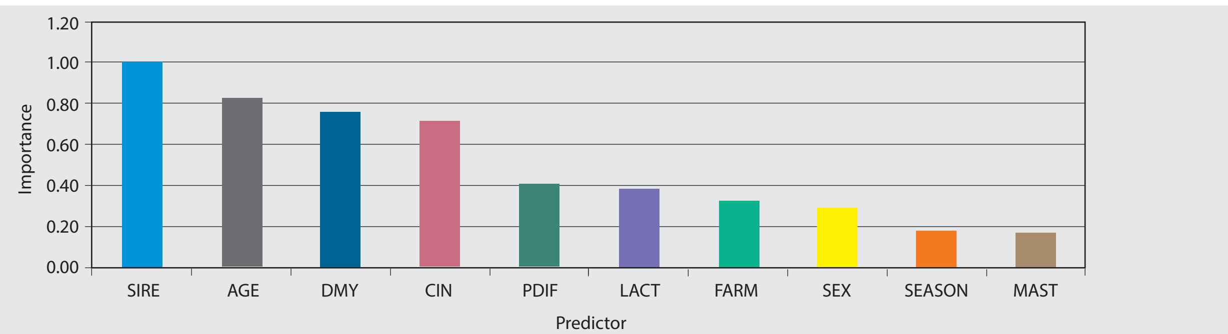

Figure 4
Predictor importance for the random forest model in cows

the percentage of properly identified easy and moderate calving events in heifers in our work was considerably lower (about 40 and 57 %, respectively). Direct comparisons can be much easier if the model with only two classes of calving difficulty is used. In this case, the accuracy in the present study was 81.25 %, which is similar to that reported by Arthur et al. (2000). Sensitivity and specificity for the two-class system in the present work were 70.91 % and 86.28 %, respectively, so the ability of the RF model to correctly detect difficult calving events in heifers was almost twice higher than that obtained by Arthur et al. (2000), whereas its effectiveness in indicating easy deliveries was comparable. The overall predictive performance of the RF model for cows developed in our study, and evaluated on the T set, was very similar to its quality determined on the L set (accuracy equal to 66 % and 68 %, respectively). Also in this case, the model was completely unable to detect difficult calving events in cows.

To better demonstrate the predictive performance of the generated models, their ability to correctly indicate calving events from the three distinguished categories was presented in the form of cumulative gains charts. As can be seen from Figure 1, RF for heifers most accurately detected difficult calving events. The results were somewhat worse for moderate and easy calving events. Almost opposite outcomes were obtained for the calving events in cows (Figure 2), for which the RF model most accurately indicated easy calving events, slightly less accurately moderate calving events and the least precisely difficult ones, although the difference in cumulative gains between calving categories depended on the proportion of records predicted by the model in the test set.

The last part of the present work was the determination of the most influential predictor variables affecting the class of calving difficulty. SIRE and AGE were the most important predictors for heifers. Dam's sire affects calving difficulty in a more complex way than does calf's sire. First of all, it contributes its own genes (indirectly) to the calf genotype influencing traits linked to direct calving difficulty such as calf size and birthweight. In addition, it transmits its genes to the dam genotype affecting traits associated with maternal calving difficulty such as dam's pelvic dimensions, its preparation for calving and its effect on calf size. Both genetic components (direct and maternal) are considered to be additive. Usually, direct heritability is estimated to be higher (0.03 to 0.17) than maternal heritability (0.02 to 0.12) (Vanderick et al., 2014). As far as the relation between the sire's effect and dystocia is concerned, Mee et al. (2011) found a significant relationship between sire's predicted transmitting ability for maternal calving difficulty and the probability of assisted delivery. The strength of this association depended on factors such as dam's parity and calf sex. It was higher for the cows in their first or second parity and those giving birth to male calves. The logit of the probability of an assisted calving rose by 0.17 and 0.13 per unit increase in the predicted transmitting ability for maternal effect at the first and second calving, respectively, and ranged from 0.11 to 0.14 for higher parities, whereas the respective values for male and female calves were 0.17 and 0.11.

The next important predictor variable for heifers was AGE. It is well known the largest difference in calving difficulty is found between heifers and higher-parity cows (Atashi et al., 2012), although in a study by Kumar et al. (2017), parity did not affect dystocia rate in Frieswal cattle, which was approx. 1.2 % in the first parity and 0.5 to 1.9 % in higher parities. In different European countries, 3 to 22 % heifers suffer from dystocia (Martin-Collado et al., 2017). Animals with inappropriate calving age (both too young and too old) and body condition at calving are more prone to dystocia than their herdmates with the optimum values of these traits (Ettema and Santos, 2004). Simerl et al. (1991) investigating five dairy breeds (mainly Holstein and Jersey) in the USA found that dystocia was more frequent in the young (less than 24 months of age) and old (more than 27 months of age) heifers, thus indicating a curvilinear relationship between age at calving and dystocia incidence. In general, the optimal age at first calving in dairy heifers is 22 to 27 months (approx. 24 months, on average) (Mee et al., 2014), although Berry and Cromie (2009), investigating seasonally calving Holstein-Friesian heifers in Ireland, proposed 25 to 27 months as an adequate age with respect to calving ease. In their study, 22-month-old heifers had a higher risk of calving assistance than 24-month-old ones, whereas heifers calving at 25 to 27 and 35 months of age had a lower risk of such assistance compared with the animals calving at 24 months of age. Also, the work by Moussavi and Mesgaran (2008) on the age at first calving in Iranian Holsteins reflected that its mean value was 27.2 months and it significantly influenced the risk of dystocia, which decreased with an increasing calving age. An increased probability of difficult calving events in young animals (less than 24 months of age), especially giving birth to male calves, is mainly attributable to insufficient pelvic area since calf size in these cases does not vary much between calving events (Mee et al., 2014). On the other hand, dystocia in Holstein heifers was also related to the higher age at first calving (2 vs. 3 years old) but this relationship resulted from an excessive body condition score at calving associated with growth potential and sexual precocity in Holsteins (Cutullic et al., 2009). As indicated by Thompson et al. (1981), there is a large genetic correlation (equal to approx. 0.84) between dystocia at first and subsequent parturitions. Moreover, dam's age at first parturition markedly affected genetic correlations between direct and maternal effects for calving difficulty in Holstein heifers (Hickey et al., 2007). Finally, it should be mentioned that some authors (Bazzi, 2010; Yıldız et al., 2011) did not report any significant relationship between calving age and difficulty.

In the case of cows, two more predictor variables were also quite influential in determining calving difficulty class. The first one was DMY; however, its effect on the risk of dystocia has not been clearly confirmed so far. In the review by Ingvartsen et al. (2003), no significant relationship between previous milk yield and an increased incidence of difficult calving events was reported, although this conclusion was only found based on three studies. But according to the cited authors, an elevated dystocia risk in cows of high milk yield seems unlikely. However, at the genetic level, a recent study by Salimi et al. (2017) on Iranian Holsteins revealed favorable but differing in magnitude (from -0.99 to -0.20) genetic correlations between dystocia score and 305-day adjusted milk yield for direct effects. The second important variable for cows was CIN. In the study by Fiedlerová et al. (2008), a statistically significant linear relationship was observed between the length of preceding calving interval and the probability of dystocia at subsequent parturition. It raised with an increasing CIN but, as reported by the authors, it could be markedly reduced by appropriate mating and avoidance of delayed services. In most other studies, an opposite relation is frequently described, i.e. a negative effect of dystocia on fertility (McHugh et al., 2011). Again, at the genetic level, Salimi et al. (2017) found unfavorable but low (approx. -0.06) genetic correlation between dystocia and CIN for the direct effects in Holsteins, whereas Muir et al. (2004) observed a positive genetic correlation (approx. 0.21) between heifer's calving difficulty and CIN between the first and second parturition. Also, small negative genetic correlations (equal to -0.03 to -0.05) were obtained for direct effects between calving difficulty and the number of days open, which is associated with CIN (Lee et al., 2003; Salimi et al., 2017).

5 Conclusions

The random forest model developed in the present study was characterized by a high percentage of correctly detected difficult calving events in heifers. However, it was completely unable to correctly detect dystocia in cows. Therefore, further improvement of its predictive performance (especially in terms of easy and moderate category in heifers and difficult category in cows) is necessary. This can be achieved by the use of more influential predictor variables in future research. The most important predictors of calving difficulty in heifers identified in the present study were sires' rank and calving age, whereas in cows these additionally included daily milk yield and the length of calving interval.

Acknowledgments

This work was supported by the Polish Ministry of Science and Higher Education [grant number 517-01-028-3962/17].

References

Abell KM, Theurer ME, Larson RL, White BJ, Hardin DK, Randle RF (2017) Predicting bull behavior events in a multiple-sire pasture with video analysis, accelerometers, and classification algorithms. Comput Electron Agric 136:221-227

Aguiar GFM, Batista BL, Rodrigues JL, Silva LRS, Campiglia AD, Barbosa RM, Barbosa F (2012) Determination of trace elements in bovine semen samples by inductively coupled plasma mass spectrometry and data mining techniques for identification of bovine class. J Dairy Sci 95(12):7066-7073

Arthur PF, Archer JA, Melville GJ (2000) Factors influencing dystocia and prediction of dystocia in Angus heifers selected for yearling growth rate. Aust J Agric Res 51(1):147-154

Atashi H, Zamiri MJ, Sayadnejad MB (2012) The effect of maternal inbreeding on incidence of twinning, dystocia and stillbirth in Holstein cows of Iran. Iran J Vet Res 13(2):93–99

Bazzi H (2010) Evaluation of non-genetic factors affecting birth weight in Sistani cattle. J Anim Vet Adv 10(23):3095-3599

Berry DP, Cromie AR (2009) Associations between age at first calving and subsequent performance in Irish spring calving Holstein–Friesian dairy cows. Livest Sci 123(1):44-54

Borchers MR, Chang YM, Proudfoot KL, Wadsworth BA, Stone AE, Bewley JM (2017) Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. J Dairy Sci 100(7):5664-5674

Breiman L (2001) Random forests. Mach Learn 45(1):5-32

Brieuc MS, Ono K, Drinan DP, Naish KA (2015) Integration of random forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). Mol Ecol 24(11):2729-2746

Burez J, Van den Poel D (2009) Handling class imbalance in customer churn prediction. Expert Syst Appl 36(3):4626-4636

Cutullic E, Delaby L, Causeur D, Michel G, Disenhaus C (2009) Hierarchy of factors affecting behavioural signs used for oestrus detection of Holstein and Normande dairy cows in a seasonal calving system. Anim Reprod Sci 113(1):22-37

Dolecheck KA, Silvia WJ, Heersche G, Chang YM, Ray DL, Stone AE, Wadsworth BA, Bewley JM (2015) Behavioral and physiological changes around estrus events identified using multiple automated monitoring technologies. J Dairy Sci 98(12):8723-8731

Ettema JF, Santos JEP (2004) Impact of age at calving on lactation, reproduction, health, and income in first-parity Holsteins on commercial farms. J Dairy Sci 87(8):2730-2742

Fiedlerová M, Rehák D, Vacek M, Volek J, Fiedler J, Simecek P, Masata O, Jílek F (2008) Analysis of non-genetic factors affecting calving difficulty in the Czech Holstein population. Czech J Anim Sci 53:284-291

Gaddis KP, Cole JB, Clay JS, Maltecca C (2016) Benchmarking dairy herd health status using routinely recorded herd summary data. J Dairy Sci 99(2):1298-1314

Grossmann E, Ohmann J, Kagan J, May H, Gregory M (2010) Mapping ecological systems with a random forest model : tradeoffs between errors and bias. Gap Anal Bull 17(1):16-22

Ha K, Cho S, MacLachlan D (2005) Response models based on bagging neural networks. J Interact Marketing 19(1):17-30

Hickey JM, Keane MG, Kenny DA, Cromie AR, Amer PR, Veerkamp RF (2007) Heterogeneity of genetic parameters for calving difficulty in Holstein heifers in Ireland. J Dairy Sci 90(8):3900-3908

Homburger H, Schneider MK, Hilfiker S, Lüscher A (2014) Inferring behavioral states of grazing livestock from high-frequency position data alone. PloS One 9(12):e114522

Ingvartsen KL, Dewhurst RJ, Friggens NC (2003) On the relationship between lactational performance and health: is it yield or metabolic imbalance that cause production diseases in dairy cattle? A position paper. Livest Prod Sci 83(2):277-308

Johnson SK, Deutscher GH, Parkhurst A (1988) Relationships of pelvic structure, body measurements, pelvic area and calving difficulty. J Anim Sci 66:1081-1088

Kovács L, Kézér FL, Ruff F, Szenci O (2017) Rumination time and reticuloruminal temperature as possible predictors of dystocia in dairy cows. J Dairy Sci 100(2):1568-1579

Kumar A, Singh U, Alex R, Kumar S (2017) Risk factors associated with abnormal parturition and their effects on production and reproductive performance in Frieswal cattle of India. Indian J Anim Res 51(2):242-246

Lee DH, Han KJ, Park BH (2003) Genetic relationship between milk yields, calving ease and days open at first lactation of Holstein cows in Korea. Interbull Bull 31:122-129

Machado G, Mendoza MR, Corbellini LG (2015) What variables are important in predicting bovine viral diarrhea virus? A random forest approach. Vet Res 46(1):85-99

Martin-Collado D, Hely F, Byrne TJ, Evans R, Cromie AR, Amer PR (2017) Farmer views on calving difficulty consequences on dairy and beef farms. Animal 11(2):318-326

McHugh N, Kearney JF, Berry DP (2011) The effect of dystocia on subsequent performance in dairy cows. Moorepark Res Rep 2011:15

Mee JF, Berry DP, Cromie AR (2011) Risk factors for calving assistance and dystocia in pasture-based Holstein–Friesian heifers and cows in Ireland. Vet J 187(2):189-194

Mee JF, Sánchez-Miguel C, Doherty M (2014) Influence of modifiable risk factors on the incidence of stillbirth/perinatal mortality in dairy cattle. Vet J 199(1):19-23

Menze BH, Petrich W, Hamprecht FA (2007) Multivariate feature selection and hierarchical classification for infrared spectroscopy : serum-based detection of bovine spongiform encephalopathy. Anal Bioanal Chem 387(5):1801-1807

Miedema HM, Cockram MS, Dwyer CM, Macrae AI (2011) Behavioural predictors of the start of normal and dystocic calving in dairy cows and heifers. Appl Anim Behav Sci 132(1):14-19

Moussavi AH, Mesgaran MD (2008) Impact of age at first calving on lactation and reproduction of first-parity Iranian Holsteins dairy cows. J Anim Vet Adv 7(2):190-195

Muir BL, Fatehi J, Schaeffer LR (2004) Genetic relationships between persistency and reproductive performance in first-lactation Canadian Holsteins. J Dairy Sci 87(9):3029-3037

Naidoo L, Cho MA, Mathieu R, Asner G (2012) Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. ISPRS J Photogramm Remote Sens 69:167-179

Nisbet R, Elder J, Miner G (2009) Handbook of statistical analysis and data mining applications. Amsterdam : Acad Pr, Elsevier, 825p

Piwczyński D, Nogalski Z, Sitkowska B (2013) Statistical modeling of calving ease and stillbirths in dairy cattle using the classification tree technique. Livest Sci 154(1-3):19-27

Rutten CJ, 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

Salimi MH, Hossein-Zadeh NG, Shadparvar AA, Eghbal AR (2017) Genetic evaluation of dystocia and its relationship with productive and reproductive traits in Holstein cows. Rev Colomb Cienc Pecuarias 30:126

Selemetas N, Ducheyne E, Phelan P, O'Kiely P, Hendrickx G, de Waal T (2015) Spatial analysis and risk mapping of *Fasciola hepatica* infection in dairy herds in Ireland. Geospat Health 9(2):281-291

Simerl NA, Wilcox CJ, Thatcher WW, Martin FG (1991) Prepartum and peripartum reproductive performance of dairy heifers freshening at young ages. J Dairy Sci 74(5):1724-1729

Thompson JR, Freeman AE, Berger PJ (1981) Age of dam and maternal effects for dystocia in Holsteins. J Dairy Sci 64(7):1603-1609

Vanderick S, Troch T, Gillon A, Glorieux G, Gengler N (2014) Genetic parameters for direct and maternal calving ease in Walloon dairy cattle based on linear and threshold models. J Anim Breed Genet 131(6):513-521

Yao C, Spurlock DM, Armentano LE, Page CD, VandeHaar MJ, Bickhart DM, Weigel KA (2013) Random forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. J Dairy Sci 96(10):6716-6729

Yıldız H, Saat N, Simsek H (2011) An investigation on body condition score, body weight, calf weight and hematological profile in crossbred dairy cows suffering from dystocia. Pak Vet J 31:125-128

Zaborski D, Proskura WS, Grzesiak W (2016) Classification of calving difficulty scores using different types of decision trees. Acta Sci Pol Zootech 15(4):55-70

Zaborski D, Proskura WS, Grzesiak W (2017) Comparison between data mining methods to assess calving difficulty in cattle. Rev Colomb Cienc Pecuarias 30(3):196-208