A comparison of 4 predictive models of calving assistance and difficulty in dairy heifers and cows

Caroline Fenlon,*[1] Luke O'Grady,† John F. Mee,‡ Stephen T. Butler,‡ Michael L. Doherty,† and John Dunnion*
*School of Computer Science, and
†School of Veterinary Medicine, University College Dublin, Belfield, D04 W6F6, Dublin 4, Ireland
‡Animal and Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, P61 P302, County Cork, Ireland

ABSTRACT

The aim of this study was to build and compare predictive models of calving difficulty in dairy heifers and cows for the purpose of decision support and simulation modeling. Models to predict 3 levels of calving difficulty (unassisted, slight assistance, and considerable or veterinary assistance) were created using 4 machine learning techniques: multinomial regression, decision trees, random forests, and neural networks. The data used were sourced from 2,076 calving records in 10 Irish dairy herds. In total, 19.9 and 5.9% of calving events required slight assistance and considerable or veterinary assistance, respectively. Variables related to parity, genetics, BCS, breed, previous calving, and reproductive events and the calf were included in the analysis. Based on a stepwise regression modeling process, the variables included in the models were the dam's direct and maternal calving difficulty predicted transmitting abilities (PTA), BCS at calving, parity; calving assistance or difficulty at the previous calving; proportion of Holstein breed; sire breed; sire direct calving difficulty PTA; twinning; and 2-way interactions between calving BCS and previous calving difficulty and the direct calving difficulty PTA of dam and sire. The models were built using bootstrapping procedures on 70% of the data set. The held-back 30% of the data was used to evaluate the predictive performance of the models in terms of discrimination and calibration. The decision tree and random forest models omitted the effect of twinning and included only subsets of sire breeds. Only multinomial regression and neural networks explicitly included the modeled interactions. Calving BCS, calving difficulty PTA, and previous calving assistance ranked as highly important variables for all 4 models. The area under the receiver operating characteristic curve (ranging from 0.64 to 0.79) indicates that all of the models had good overall discriminatory power. The neural network and multinomial regression models performed best, correctly classifying 75% of calving cases and showing superior calibration, with an average error in predicted probability of 3.7 and 4.5%, respectively. The neural network and multinomial regression models developed are both suitable for use in decision-support and simulation modeling.
Key words: dystocia, calving assistance, machine learning, external evaluation

INTRODUCTION

Difficult parturition (dystocia) has severe consequences for the welfare of both the dam and calf, including pain, increased risk of surgery, morbidity linked to other diseases, mortality, and culling (Huxley and Whay, 2006; Mee, 2008a). Dystocia also has significant direct and indirect economic consequences (Dematawewa and Berger, 1997). Direct costs include veterinary assistance/treatment at the point of calving and increased risk of dam and calf mortality (Ettema and Santos, 2004). Indirect costs include treatment of dystocia-associated transition cow disorders (Erb et al., 1985; LeBlanc, 2012) and sequelae such as losses in genetic gains associated with mortality or culling, increased BCS loss, reduced milk production, and impaired reproductive performance (Fourichon et al., 1999, 2000; Berry et al., 2007; Fenlon et al., 2017).

The classification systems for calving difficulty vary internationally (Mee, 2008b), but generally consist of unassisted calving (**UC**) events (unobserved or observed), unrequired/nonessential assistance, and required/essential interventions. In Ireland, calving difficulty is recorded by farmers using an ordinal scale from 1 to 4 (Mee et al., 2011). In dairy cattle, dystocia generally occurs in less than 5% of calving events (Mee, 2008b), though this figure varies by country and greatly between herds. The most recently reported incidence of scores 1 to 4 in Irish Holstein-Friesian cows were 68.9, 24.3, 4.3, and 2.5%, respectively (Mee et al., 2011)

Risk factors for dystocia associated with the dam have been identified: parity, with first parity animals having a higher incidence of calving difficulty (Johanson and Berger, 2003; Mee et al., 2011); nutritional status and its interaction with age at first calving, particularly undercondition or overcondition at calving (Drew, 1986); and previous history of dystocia (Mee et al., 2011). Similarly, risk factors associated with the calf have been identified: higher birth weight and higher weight relative to the dam's BW (Berry et al., 2007; Mee, 2008b), gestation length, plurality (Ettema and Santos, 2004), and sex (Ettema and Santos, 2004). Embryos produced in vitro resulted in calves with longer gestation, greater birth weight, and greater levels of difficulty than calves arising from AI or in vivo embryos (Kruip and Den Daas, 1997).

Accurate prediction of the risk of dystocia at an individual cow level could inform preventative actions and farm management decisions. A limited number of studies have attempted to predict the probability of dystocia. Binary models comparing assisted calving and UC events in various breeds of dairy and beef cows have been created using logistic regression (Johanson and Berger, 2003; Bureš et al., 2008; Mee et al., 2011). A study of Canadian Holstein calving events performed least squares analysis to model 4 levels of calving difficulty (Klassen et al., 1990). However, few studies have evaluated the predictive ability of calving difficulty models. A series of studies used classification trees, support vector machines, neural networks (**NN**), and generalized linear models to distinguish between UC and difficult calving (**DC**) events for Polish Holstein-Friesians managed in indoor production systems (Zaborski and Grzesiak, 2011; Zaborski et al., 2014a,b). The models were evaluated using root mean square error and prediction sensitivity, specificity, and accuracy. Another study of Polish Holstein-Friesian cows modeled 4 levels of calving ease using classification trees and again evaluated the results using classification error measures (Piwczyński et al., 2013).These discrimination tests measure a model's ability to correctly classify cases, (i.e., the separation between the possible outcomes). Another type of evaluation is available for probabilistic models (Tedeschi, 2006). Methods of calibration allow the identification of any areas of poor fit or bias in the predictions, by comparing predicted probabilities to true proportions of events in groups of similar records.

The aim of the present study was to create and evaluate predictive models of calving assistance and dystocia for use in decision-support tools and simulation modeling. Our objective was to employ a range of machine learning techniques to create a model using data commonly recorded by dairy farmers.

MATERIALS AND METHODS

Data

A total of 1,686 records of calving events from the years 2000 to 2010 were sourced from the Ballydague and Curtins research herds at Teagasc's Animal and Grassland Research and Innovation Centre, Moorepark, County Cork, Ireland. These herds were representative of recommended Irish grass-based farming systems with a diverse range of cow genetics (Horan et al., 2005). An additional 390 calving events from 2015 were available from 8 herds of Holstein-Friesian cows. These herds were involved in a herd health and fertility consultancy program operated by the School of Veterinary Medicine, University College Dublin (Somers et al., 2015). All of the herds operated seasonal breeding, with the recorded calving events happening between January and April.

Along with the dam tag number and calving date, records available from each herd's herd management software included calving difficulty score and BCS. Calf-level information included birth weight, sex, sire, and binary indicators of stillbirth (dead calf born at term) and twinning. Details of AI service dates, pregnancy diagnosis, and BCS from the preceding lactation were also available. Additionally, breed, genetic economic breeding values and PTA for traits of economic importance were available from the Irish Cattle Breeders Federation (**ICBF**) national database. In total, approximately 120 variables were available for consideration in the modeling process.

The date of each service event that resulted in conception was confirmed by ultrasound scan between 30 and 60 d postservice or by subsequent calving 282 ± 15 d postservice. The gestation length was calculated from this service event.

Calving Assistance

Calving assistance was measured on an ordinal scale: 1 = UC; 2 = slight assistance (**SA**; assistance by 1 person with no use of mechanical calving aids); 3 = considerable difficulty (**CD**; with intervention by 2 people or the use of mechanical calving aids); 4 = veterinary assistance (**VA**; with or without the need for surgical intervention).

As model outcomes, SA was retained as level 2, and a DC was defined as levels 3 and 4 combined as a single group. Where twinning occurred and different calving difficulty scores were recorded for each calf, the higher value was used as the overall score for the calving. Three additional binary variables were created indicating pre-

vious calving difficulty. One related to whether a cow had been given any previous calving assistance score ≥2 and another to any calving assistance score ≥3. The third variable indicated a score ≥2 at the previous calving only (**PrevCalvAss**). Heifers were coded as having no previous calving assistance for all 3 variables.

Twinning

The birth of multiple calves was entered in the data as a binary variable. An additional binary variable denoting previous occurrences of twins (defaulting to false for first calving) was created. The analysis did not consider ultrasound scans indicating the presence of multiple embryos that did not result in the birth of multiple calves. Twin births were analyzed as single calving events. Where a twin birth occurred, the weights of the 2 calves were summed to give a total calving birth weight.

Calf Sex

The birth of at least one male calf (including those born with a female twin) was indicated with a binary variable.

Cow BCS

The BCS at conceiving service and at calving were considered in the analysis. As the number of records close to the conceiving service was low, BCS was instead estimated using splines based on all of the cow's measurements during the lactation. Where BCS was not available for the day of calving, the closest measurement no more than 20 d before calving was used. If the animal was not scored in this time range, the highest of the scores taken up to 33 d before calving and 10 d after calving was used. In addition, the nadir BCS during the lactation before the calving under analysis was calculated. Service and nadir BCS scores were not available for maiden heifers. All BCS measurements were considered as both continuous variables and discrete scores, by rounding to the nearest 0.25 unit of BCS and grouping values with few records.

Genetics and Breed

The PTA values for all traits used in the calculation of the Irish national breeding program's economic breeding index (**EBI**) values were sourced in July 2014 from the ICBF national database for all cows in the data set. These included the overall EBI and the sub-indexes for milk production, fertility, calving, beef, maintenance, and health, as well as PTA values for progeny carcass weight, progeny carcass conformation, progeny carcass fat score, cull cow carcass weight, cow weight, direct calving difficulty (**DCD**), gestation length, calf mortality, maternal calving difficulty (**MCD**), milk yield (kg), milk fat percentage, milk protein percentage, milk fat yield (kg), milk protein yield (kg), SCS, locomotion score, calving interval, and survival. For each dam, the proportions of the 2 predominant breeds were also available from the ICBF national database. Breed and DCD were also available for the sire of the calf or calves born at each calving.

Model-Building Procedure

All analyses were carried out using the R statistical programming language (R Core Team, 2013, R Foundation for Statistical Computing, Vienna, Austria).

An initial univariate analysis using multinomial regression (**MR**) was conducted to screen candidate variables for inclusion in the multivariate models and to test for nonnormal relationships with dependent likelihoods. The univariate relationship between each of the explanatory variables with the probabilities of calving assistance and dystocia was estimated. As random effects were not possible for categorical outcomes in the modeling algorithms chosen, repeated and unmeasurable effects (cow, herd, month, and year) were also eligible for inclusion as fixed effects.

The calving difficulty models were built using the bootstrapping resampling technique [from the R "caret" library (Kuhn et al., 2016)], with 500 iterations. A randomly selected 70% of the available data (1,454 calving events) was used to train the bootstrapped models, with the other 622 calving events held back to evaluate the predictive ability of the models. The "createDataPartition" function of the caret library was used to retain an equal distribution of the calving difficulty scores across the 2 data sets.

Four machine learning algorithms were used in the analysis of the 3 levels of calving difficulty (UC, SA, and DC). These are popular techniques for the classification of categorical values and all are suitable for use with bootstrapping. All of the techniques are capable of predicting cumulative probabilities for categorical outcomes, as well as the most likely outcome value.

Multinomial Regression

Similarly to binary logistic regression, MR relates the independent variables to the probability of each category in the dependent variable (Venables and Ripley, 2013). Continuous values are assumed to have a normal

relationship with the resultant probabilities. The probabilities for levels of categorical predictor values are estimated individually.

Forward stepwise modeling (R "add1" function) with MR was used to identify the variables and interactions to be included in this study's models. At each step, the most significant biologically relevant variable with $P \leq 0.05$ was added. The model was refined by removing variables not significant at $P \leq 0.20$ (R "drop1" function). This testing of variables was repeated until no more factors were available to improve the model. Once no more individual variables could be included or excluded from the model, all 2-way interactions between the variables in the fixed effects model were tested.

Decision Tree

The classification and regression decision tree (**DT**) method based on the work of Breiman et al., (1984) was used (Therneau et al., 2015). Each node in the tree consists of a binary splitting rule, resulting in the potential for multiple splits in categorical or continuous independent variables. Continuous values are automatically split at the optimal point. Explicit interactions are not permitted in the definition of the model in this implementation, but may be implicitly represented within the resultant tree branches. The classification and regression tree algorithm uses the Gini index to determine the inclusion and order of the variables (Wu et al., 2007); the variables included in the model definition are not necessarily all included in the final DT.

Random Forest

Random forests (**RF**) are an ensemble DT method (Hothorn et al., 2006). The algorithm uses internal bootstrapping with random feature selection to train several trees. As with the standard DT, interactions are not possible. The number of bootstrapping aggregations is required as a user input. In this study, 500 trees with the number of variables used ranging from 2 to 10 were created throughout the bootstrapping procedure.

Neural Network

Artificial NN are models inspired by the workings of real neurons. They consist of one or more layers of nodes which can model complex relationships between dependent and independent variables. The implementation used in this study created a feed-forward NN with a single hidden layer (Venables and Ripley, 2013). The networks were trained using the least squares method with a maximum of 100 iterations. The number of nodes within the network can be altered by the user. During the model training, multiple different-sized networks were built, ranging from 5 to 20 nodes, each with the bootstrapping process described above. The network with the optimal accuracy was chosen as the final model.

Variable Importance Ranking

The P-values for each variable or category in the final bootstrapped model were calculated for the MR. Variable rankings for the other models were calculated using the "varImp" function of the caret library. This calculates variable importance in an appropriate way for each of the different machine learning methods. Variables in the DT are ranked by the number of decision nodes in which they were contained. The importance function for the RF measures the change in accuracy caused by the inclusion of each variable. The magnitudes of the coefficients contributing to the prediction are calculated for the NN.

Model Evaluation

As the aim of this study was to identify the best predictive model of calving difficulty, the models were evaluated in terms of their predictive ability. As explained earlier, when measuring the accuracy of probabilistic predictions, 2 forms of evaluation are possible: discrimination and calibration. In the case of categorical outcome values such as the one modeled here, each potential level must be evaluated separately. For each difficulty level (UC, SA, and DC), binary comparisons were made between group membership and nonmembership.

Measures of discrimination evaluate the ability of the models to distinguish between different difficulty levels. Classification matrices for each model were tabulated from the most likely outcomes. Sensitivity (true-positive rate), specificity (true-negative rate), positive predictive value (**PPV**; the proportion of correctly predicted positive outcomes), negative predictive value (**NPV**; the proportion of correctly predicted negative outcomes) and accuracy (the proportion of correctly predicted outcomes) were calculated. The F-measure is the harmonic mean of sensitivity and PPV and is used as a measure of overall discrimination ability (van Rijsbergen, 1979). For each model, a receiver operating characteristic (**ROC**) curve was created by plotting the tradeoff between the false-positive and true-positive rates as the discrimination threshold is altered ["ROCR" R package (Sing et al., 2005)]. This was carried out using a one-versus-all approach for each calving difficulty

level (Hand and Till, 2001). The area under the ROC curve (**AUC**) indicates the probability that the model will predict a higher probability for a randomly chosen positive instance than for a randomly chosen negative instance. A separate AUC was calculated for each of the 3 modeled difficulty levels and an overall average was determined from these. To create a baseline for comparison to the model outputs, predictions were also calculated using the population average occurrence rates for the 3 calving difficulty levels, equivalent to random selection. To allow comparison with previous studies the measures of accuracy were also calculated using 2 alternative classifications of calving assistance: (1) UC versus the combined group of SA and DC scores (SA/DC), and (2) the combined group of UC and SA scores (UC/SA) versus DC.

Rather than assessing how well a model assigns predictions to the correct category, calibration techniques test the accuracy of the predicted category membership probability (i.e., determining if the observed frequency of a difficulty level is similar to the predicted probability) within groups of similar records (Hosmer et al., 2013). This can identify bias in the predicted probabilities. Calibration plots were created using 10 equidistant probability groups, with F-distribution confidence intervals ["PresenceAbsence" R package (Freeman and Moisen, 2008)]. Deviance values were calculated as the difference between the actual outcome (coded as 1 or 0) and the predicted probability of that outcome and were arranged into groups of equal size sorted by predicted probability ["arm" R package (Gelman and Su, 2015)]. The absolute values of the deviances were averaged to find the mean absolute calibration error (**MACE**), which indicates the average error of the predicted probabilities for the category.

RESULTS

The training data included 1,078 (74.1%) UC, 289 (19.9%) SA, and 87 (5.9%) DC calving events. The testing data had the same distribution of difficulty levels. Three and a half percent of the calving events resulted in multiple births, and 2.17% resulted in a stillborn calf. Across all years, the proportions of cows that were in parity 1, 2, and ≥3 were 34.49, 25.48, and 40.03%, respectively. Additional descriptive statistics, displayed by herd, are summarized in Table 1.

Univariate Multinomial Regression Analysis

In the univariate analysis, lower BCS loss during the preceding lactation resulted in a lower risk of calving difficulty. This was also the case for animals with a higher nadir BCS. When BCS nadir was later in lacta-

Table 1. Descriptive statistics by herd (mean, with SD in parentheses where appropriate)

Item[1]	Research herds: 10 yr		Commercial herds: 1 yr							
	1	2	1	2	3	4	5	6	7	8
Parity	2.54 (1.67)	2.02 (1.13)	2.74 (1.71)	2.85 (2.00)	2.63 (1.55)	3.44 (2.50)	2.64 (1.77)	1.88 (1.17)	3.69 (2.05)	3.04 (1.97)
Herd size	128.30 (49.89)	35.60 (22.63)	62	33	73	27	45	25	58	67
Calving BCS	3.34 (0.36)	3.22 (0.36)	3.13 (0.21)	3.16 (0.21)	3.00 (0.08)	3.09 (0.17)	3.17 (0.22)	3.22 (0.17)	3.22 (0.09)	3.39 (0.20)
Parity 1	3.49 (0.35)	3.39 (0.37)	3.15 (0.13)	3.28 (0.08)	3.01 (0.05)	3.30 (0.11)	3.23 (0.24)	3.25 (0.00)	3.25 (0.00)	3.49 (0.10)
Parity 2	3.18 (0.28)	3.07 (0.30)	3.10 (0.15)	3.03 (0.15)	2.98 (0.08)	3.00 (0.20)	3.10 (0.22)	3.16 (0.19)	3.21 (0.09)	3.38 (0.21)
Parity ≥3	3.31 (0.37)	3.14 (0.31)	3.13 (0.27)	3.17 (0.24)	3.00 (0.09)	3.07 (0.11)	3.17 (0.21)	3.25 (0.31)	3.21 (0.09)	3.34 (0.22)
Calving events (%)										
January	24	25	0	27	1	33	29	28	0	31
February	49	56	92	61	66	67	64	68	26	34
March	19	17	6	12	27	0	7	4	47	18
April	8	2	2	0	5	0	0	0	28	16
May	1	0	0	0	0	0	0	0	0	0
DCD PTA	2.76 (1.40)	2.71 (1.53)	2.13 (0.58)	2.75 (0.57)	2.66 (0.55)	2.51 (0.65)	2.52 (0.74)	2.18 (0.41)	2.48 (0.83)	3.10 (0.93)
MCD PTA	6.74 (1.20)	7.20 (1.30)	6.53 (0.58)	6.48 (0.61)	6.56 (0.64)	6.60 (0.92)	6.26 (0.65)	6.30 (0.80)	6.67 (0.69)	6.43 (0.91)
Sire DCD PTA	2.41 (1.89)	2.40 (1.84)	1.06 (0.78)	1.66 (0.50)	2.31 (1.17)	2.55 (0.87)	1.94 (1.02)	1.88 (0.49)	2.99 (1.56)	2.54 (0.58)
Calving type (%)										
Unassisted calving	888 (67)	291 (80)	61 (98)	32 (97)	71 (97)	27 (100)	38 (84)	24 (96)	53 (91)	55 (82)
Slight assistance	345 (26)	44 (12)	0 (0)	1 (3)	2 (3)	0 (0)	6 (13)	1 (4)	5 (9)	8 (12)
Difficult calving	88 (7)	30 (8)	1 (2)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Total calf weight (kg)	38.57 (6.76)	37.77 (7.99)	35.08 (8.51)	39.12 (5.40)	39.02 (4.68)	43.15 (7.98)	35.14 (5.30)	35.56 (5.83)	42.14 (7.05)	41.61 (6.69)

[1]DCD PTA = direct calving difficulty PTA (% of difficult calving events); MCD PTA = maternal calving difficulty PTA (% of difficult calving events).

tion, this very slightly increased the probability of a cow requiring SA.

Month of calving did not affect calving difficulty. Twin calves resulted in a higher probability of any assistance being required. Previous occurrences of twinning did not affect future calving difficulty. The birth of one or more male calves increased the likelihood of needing assistance and of dystocia by 1.4- and 2.2-fold, respectively.

Dystocia at any prior calving event significantly increased the probability of requiring SA. Assistance at the most recent previous calving (PrevCalvAss) resulted in more than twice the likelihood of assistance at a subsequent calving compared with unassisted animals.

Higher sire DCD resulted in a significantly 1.2 times higher probability of SA and 1.4 times higher probability of DC. The sire's breed was a significant effect; when compared with Aberdeen Angus sires, Belgian Blue and Montbeliarde sires resulted in a higher likelihood of DC and Friesian and Holstein sires resulted in greater probability of SA. Jersey-sired calves had a lower probability of experiencing either SA or DC calving events, and Norwegian Red sires resulted in a lower probability of SA.

The greater the proportion of Holstein in the dam's breed, the higher the probability of SA. Several EBI genetic traits were significant in the univariate analysis. Higher overall EBI was associated with a lower probability of SA and DC. Higher milk fat yield and protein yield PTA slightly increased the probability of SA but lowered the probability of DC. Higher milk fat percentage and milk protein percentage PTA resulted in lower levels of both SA and DC. A longer calving interval PTA, higher calving sub-index, progeny carcass weight, cull cow carcass weight, cow weight, survival, and direct (DCD) and maternal (MCD) calving difficulty traits were associated with higher probabilities of both SA and DC.

Multivariate Models

The significant variables identified after the stepwise MR modeling procedure were dam DCD and MCD PTA, proportion of Holstein breed, (continuous) BCS at calving, parity (categorical, with ≥3 grouped), twinning, PrevCalvAss, sire DCD PTA, and sire breed. Two 2-way interactions were statistically significant: calving BCS and previous calving difficulty, and MCD with sire breed. Odds ratios from the multivariate MR model are summarized in Table 2. Some of the odds ratios (e.g., particular sire breed groups) are 0, because only UC events existed in the randomly selected training data set for these particular values. The graph of the DT model is shown in Figure 1.

Variable importance for the MR is indicated by the P-values in Table 2. Most of the values were significant in the final bootstrapped model with $P \leq 0.05$. Twinning and the proportion of Holstein were not statistically significant. The sire breeds Normande, Swiss Red, and other also had P-values >0.05. Variable rankings for the remaining 3 models are summarized in Table 3. The DT only included 2 sire breeds (Belgian Blue and Holstein). Neither classification tree method (DT and RF) included twinning. The RF did not incorporate 3 breeds: Aberdeen Angus, SR, and other. Variable rankings in DT and RF were largely the same, with no interactions explicitly modeled. The NN incorporated all of the variables and interactions that were significant in MR. Sire breed was the least important variable in all 3 models. Its interaction with MCD was also lowly ranked in NN, with the most important breed (Belgian Blue) at position 7. Calving BCS was the highest ranked variable in DT and RF and the second highest in NN. The DCD, MCD, and sire DCD were in the top 5 rankings for both DT and RF. The DCD was the most important genetic trait in NN, at position 6. Parity was important in NN, with parity 2 ranked first and parity ≥3 ranked third. Only parity ≥3 was modeled in the DT (at position 6). Parity values ranged from position 4 (≥3) to 10 (parity 2) in NN. The dam's proportion of Holstein breed was positioned seventh (DT and RF) and eighth (NN). The importance of PrevCalvAss ranked from position 4 (DT) to 9 (NN).

Model Evaluation

Results of discrimination tests performed on the test data subset are presented in Table 4. The models had overall average F-measures ranging from 39.8 to 41.7%. All of them were significantly better than the average F-measure for the random baseline model. Sensitivity and specificity varied widely by model and difficulty level. The DT and NN had 0% sensitivity for DC, because they did not correctly predict any calving events with DC. Specificity was very high for the SA and DC outcomes, meaning that few of those not needing assistance were predicted as needing assistance. In contrast, the sensitivity and F-measure for UC events were very high and specificity was low. All models showed superior performance relative to the use of random selection using herd average incidences.

Average predicted probabilities were similar across all of the models. However, the range of predicted probabilities differed across the models. Whereas MR and NN predicted UC probabilities as low as 1 and 7%, the minimum predictions from DT and RF were 21 and 19%, respectively. The MR, RF, and NN all generated probabilities less than 2% for SA, whereas

the lowest DT prediction was 10%. All 4 models predicted minimum probabilities of 2% or less for DC. The maximum probability for DC ranged from 37% (DT) to 83% (MR).

Calibration curves for the models are presented in Figure 2. The ROC curve for DC is shown in Figure 3. The MACE and AUC, calculated by difficulty level and averaged for overall figures, are presented in Table 5. The MR had the lowest MACE for the prediction of SA, at 5.0%. The DT had the lowest AUC across all 3 levels (average 0.64). The NN had the lowest MACE for the prediction of DC, at 1.7%. The DT calibration plot has one SA group point below the diagonal line, indicating a group of values with predicted probabilities higher than the true incidence of UC. The RF had one UC group below the diagonal and one SA group above the line. NN had the lowest average MACE, at 3.7%. Both MR and NN were well calibrated, with the diagonal line within the confidence intervals of all points.

DISCUSSION

Our study is one of few to predict multiple levels of calving difficulty and, to the best of our knowledge, the first to incorporate RF techniques. In addition, alongside the more commonly used discrimination tests, this is the first application of calibration techniques to the evaluation of predictive models for calving difficulty.

A key consideration for the acceptance and validity of a predictive decision-support model is the biological plausibility of the variables contained in the model. The variables found to be significant in the present study's modeling process agree with previous findings. Klassen et al. (1990), Bureš et al. (2008), and Mee et al. (2011) all included some measure of age, either by parity or primiparity. These age effects are likely to reflect some of the increased calving difficulty associated with small pelvic measurements reported by Bureš et al. (2008). Unfortunately, pelvic measurements were

Table 2. Multinomial regression model odds ratios (95% CI) and P-values[1]

Variable	Odds ratio–SA	Odds ratio–DC	P-value
DCD	1.21 (1.07–1.37)	1.72 (1.44–2.06)	<0.001
MCD	1.07 (0.64–1.76)	0.42 (0.14–1.24)	0.23
Sire breed			
ABA	1	1	Reference[2]
BB	0.00 (0.00–0.03)	0.00 (0.00–0.00)	<0.001
FR	0.08 (0.00–6.00)	0.00 (0.00–0.06)	0.02
HO	0.03 (0.00–1.56)	0.00 (0.00–0.02)	0.01
JE	0.03 (0.00–3.99)	0.00 (0.00–0.07)	0.02
MO	0.00 (0.00–3.31)	0.00 (0.00–0.21)	0.03
NO	5.80 (0.00–3.6e+07)	0.00 (0.00–2.0e+06)	0.76
NR	0.00 (0.00–4.9e+08)	0.00 (0.00–0.01)	<0.001
SR	0.54 (0.00–18,134.26)	0.43 (0.00–43,994.19)	1.00
Other	16.93 (0.00–6.7e+08)	0.00 (0.00–127.04)	0.26
Calving BCS	14.89 (8.10–27.38)	4.39 (1.72–11.24)	<0.001
Sire DCD	1.29 (1.14–1.47)	1.43 (1.19–1.72)	<0.001
Parity			
1	1	1	Reference[2]
2	0.75 (0.48–1.18)	0.37 (0.16–0.81)	0.03
≥3	0.32 (0.22–0.48)	0.26 (0.14–0.50)	<0.001
PrevCalvAss	2,958.70 (145.00–60,372.99)	191.37 (1.70–21,571.16)	<0.001
% HO	1.57 (0.85–2.91)	1.20 (0.48–2.99)	0.36
Twin calves	1.61 (0.76–3.43)	2.84 (0.99–8.14)	0.15
MCD × sire breed			
ABA	1	1	Reference[2]
BB	5.83 (1.50–22.56)	14.35 (2.79–73.86)	<0.001
FR	1.40 (0.78–2.53)	4.92 (1.46–16.57)	0.02
HO	1.60 (0.94–2.71)	4.98 (1.64–15.12)	<0.001
JE	1.43 (0.71–2.86)	5.47 (1.43–20.89)	0.03
MO	2.11 (0.70–6.33)	4.50 (1.14–17.73)	0.05
NO	0.52 (0.03–9.54)	2.56 (0.05–139.16)	0.81
NR	4,019.26 (0.11–1.4e+08)	51.20 (2.33–1,127.68)	<0.001
SR	0.18 (0.00–6.6e+25)	0.21 (0.00–1.1e+27)	0.99
Other	0.49 (0.04–6.70)	8.26 (0.57–118.61)	0.24
Calving BCS × PrevCalvAss	0.11 (0.04–0.28)	0.21 (0.05–0.88)	<0.001

[1]DCD = direct calving difficulty PTA (% of difficult calving events); MCD = maternal calving difficulty PTA (% of difficult calving events); ABA = Aberdeen Angus; BB = Belgian Blue; FR = Friesian; HO = Holstein; JE = Jersey; MO = Montbeliarde; NO = Normande; NR = Norwegian Red; SR = Swiss Red; % HO = dam's proportion of Holstein breed; PrevCalvAss = assistance or difficulty at the previous calving; SA = slight assistance; DC = difficult calving.

[2]P-values are not generated for reference values of categorical variables.

8

FENLON ET AL.
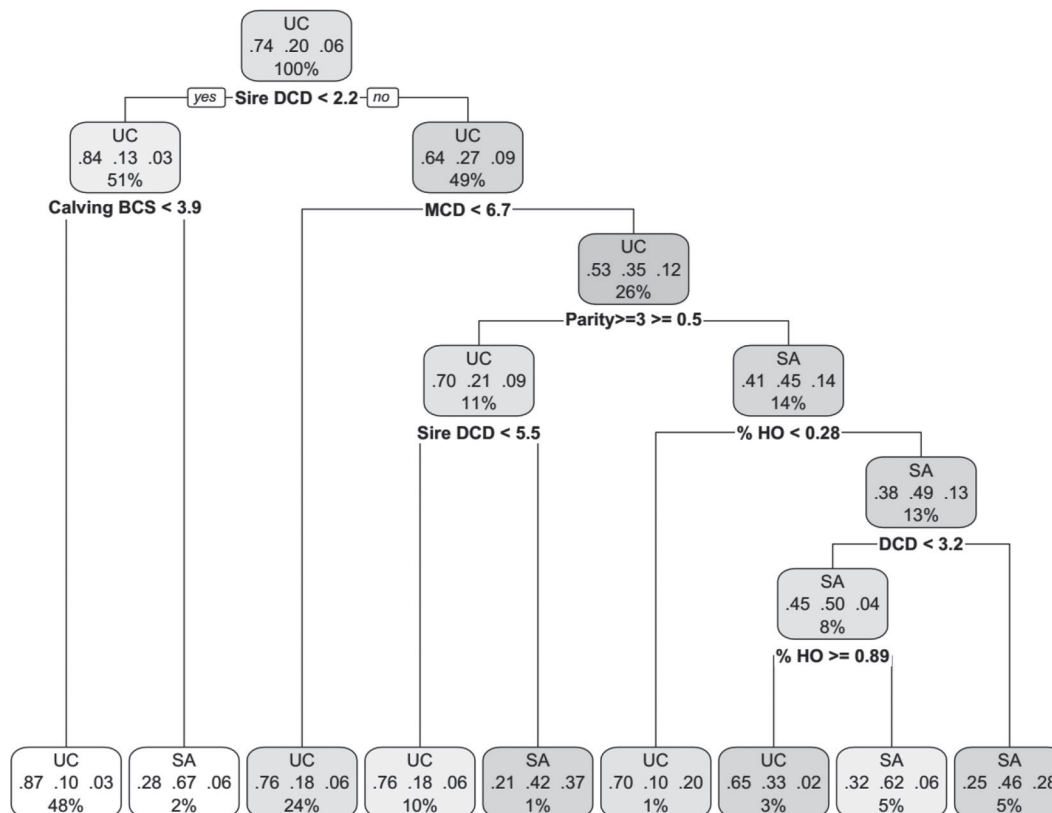
Figure 1. Graph of the decision tree model. Each node includes the likelihood (in order) for the 3 levels of calving difficulty and the proportion of training instances used to calculate the probabilities for that node. No difficult calving events were predicted by the decision tree. UC = unassisted calving; SA = slight assistance. DCD = direct calving difficulty; MCD = maternal calving difficulty; % HO = dam's proportion of Holstein breed.

not available for analysis in the present study. Dystocia at the previous calving, sire DCD, and sire breed were found to be significant by Mee et al. (2011). Klassen et al. (1990) and Mee et al. (2011) both identified twin calving events as increasing the risk of calving difficulty.

Our results differ for some of the previously reported risk factors. Greater BCS at calving resulted in greater likelihood of difficulty in all of our models. When taken as discrete values in univariate analysis, this effect was significant only when the BCS was ≥3.5; low BCS values did not differ significantly from the base value of 3.25. While Keady et al. (2001) and Berry et al. (2007) found no effect of weight or BCS on calving difficulty, this finding may correspond to the problems arising from overfeeding reported by Grunert (1979). We did not find any temporal values (year or month of calving) to be significant. As random effects were not possible using the machine learning implementations used, including temporal or herd variables would prohibit the use of the models for prediction. In contrast, Mee et al. (2011) and Zaborski and Grzesiak (2011) found that the season of calving influenced the incidence of

difficulty. However, no calving events occurred during autumn in the data used for our study.

The interactions present in our model have not previously been reported. These include the reduced effect of increased BCS in cows with a history of calving difficulty. This suggests that these are not full additive genetic effects in relation to increased calving difficulty. The interaction between MCD and sire breed suggests that some breeds compound the effect of the MCD trait.

For utility as a decision-support tool, the variables required in the model should be readily available to target users. This study's models include only variables that can easily be measured or collected by dairy farmers. In contrast with some previous studies, no measures of calf weight (Piwczyński et al., 2013) or genomic data (Zaborski and Grzesiak, 2011; Zaborski et al., 2014b) were significant in the models.

The models obtained in this study performed similarly to previous work involving analogous modeling techniques. The classification tree methods used by Piwczyński et al. (2013) resulted in AUC from 0.69 to

Table 3. Variable importance ranking for decision tree (DT), random forest (RF), and neural network (NN) machine learning models of calving difficulty

Ranking	DT	RF	NN
1	Calving BCS	Calving BCS	Parity 2
2	DCD	MCD	Calving BCS
3	MCD	Sire DCD	Parity ≥3
4	PrevCalvAss	Parity ≥3	Sire breed HO
5	Sire DCD	DCD	Calving BCS × PrevCalvAss
6	Parity ≥3	PrevCalvAss	DCD
7	% HO	% HO	MCD × sire breed BB
8	Sire breed BB	Sire breed HO	% HO
9	Sire breed HO	Parity 2	PrevCalvAss
10		Sire breed MO	Twin calves
11		Sire breed BB	MCD × sire breed MO
12		Sire breed FR	Sire breed FR
13		Sire breed NR	MCD × sire breed FR
14		Sire breed JE	MCD × sire breed NO
15		Sire breed NO	Sire breed JE
16			MCD
17			MCD × sire breed NR
18			Sire breed NR
19			Sire DCD
20			MCD × sire breed JE
21			MCD × sire breed HO
22			MCD × sire breed SR
23			MCD × sire breed other
24			Sire breed BB
25			Sire breed MO
26			Sire breed NO
27			Sire breed other
28			Sire breed SR

[1]DCD = direct calving difficulty PTA (% of difficult calving events); MCD = maternal calving difficulty PTA (% of difficult calving events); BB = Belgian Blue; FR = Friesian; HO = Holstein; JE = Jersey; MO = Montbeliarde; NO = Normande; NR = Norwegian Red; SR = Swiss Red; % HO = dam's proportion of Holstein breed; PrevCalvAss = assistance or difficulty at the previous calving.

0.71 (calculated from the same data used for training the models), somewhat higher than the average AUC for DT in our study (0.64), with the accuracy reported (61.5% correctly classified calving events) very similar to our DT model. Zaborski et al. (2014a) compared assisted calving and UC events with boosted classification

Table 4. Discrimination (%) results (sensitivity, specificity, F-measure, and accuracy) by calving difficulty level for test data predictions[1]

| Discrimination	Calving type	Random	DT	MR	RF	NN
Sensitivity	UC	74.03	92.21	94.16	96.32	93.07
	SA	20.33	34.15	26.02	20.33	27.64
	DC	5.41	0.00	2.70	2.70	0.00
Specificity	UC	26.25	34.38	30.63	23.13	30.00
	SA	79.96	90.18	92.79	94.39	91.18
	DC	94.02	100.00	98.80	100.00	99.66
F-measure	UC	74.19	85.80	86.31	86.41	85.66
	SA	20.16	39.25	33.51	28.41	33.83
	DC	5.41	0.00	4.44	5.26	0.00
	Average	33.25	41.68	41.42	40.03	39.83
Positive predictive value	UC	74.35	80.23	79.67	78.35	79.34
	SA	20.00	46.15	47.06	47.17	43.59
	DC	5.41	0.00	12.50	100.00	0.00
Negative predictive value	UC	25.93	60.44	64.47	68.52	60.00
	SA	80.28	84.75	83.57	82.78	83.64
	DC	94.02	94.05	94.14	94.20	94.03
Accuracy	UC vs. SA vs. DC	59.32	75.24	75.24	75.72	74.60
	UC vs. SA/DC	61.74	77.33	77.81	77.49	76.85
	UC/SA vs. DC	88.75	94.05	93.09	94.21	93.73

[1]DT = decision tree; MR = multinomial regression; RF = random forest; NN = neural network; UC = unassisted calving; SA = slight assistance; DC = difficult calving.

trees. The AUC values on a held-back data subset were 0.81 and 0.86 for the separate models trained for heifers and cows, respectively, and the accuracies were 89.4 and 75.1%, respectively. Zaborski and Grzesiak (2011) had AUCs ranging from 0.63 to 0.84 for various NN techniques and reported an accuracy of 80.5% for the best performing binary model. The AUC values were 0.67 and 0.80 for our DT and NN models, respectively. In all of the above studies by Zaborski, calving difficulty was predicted as a binary variable (DC = CD + VA,

as in the current study). A generalized linear model with an ordinal 3-level response correctly classified 66.33% of easy calving events and 69.01% of moderate assistance (Zaborski et al., 2016), but did not predict any DC events. This was similar to our MR results; individual accuracy values for UC and SA were 62.39 and 59.40%, respectively, whereas 50.75% of DC events were correctly predicted. The most similar NN model (Zaborski and Grzesiak, 2011) had a sensitivity of 80% (for DC events), which was lower than the performance
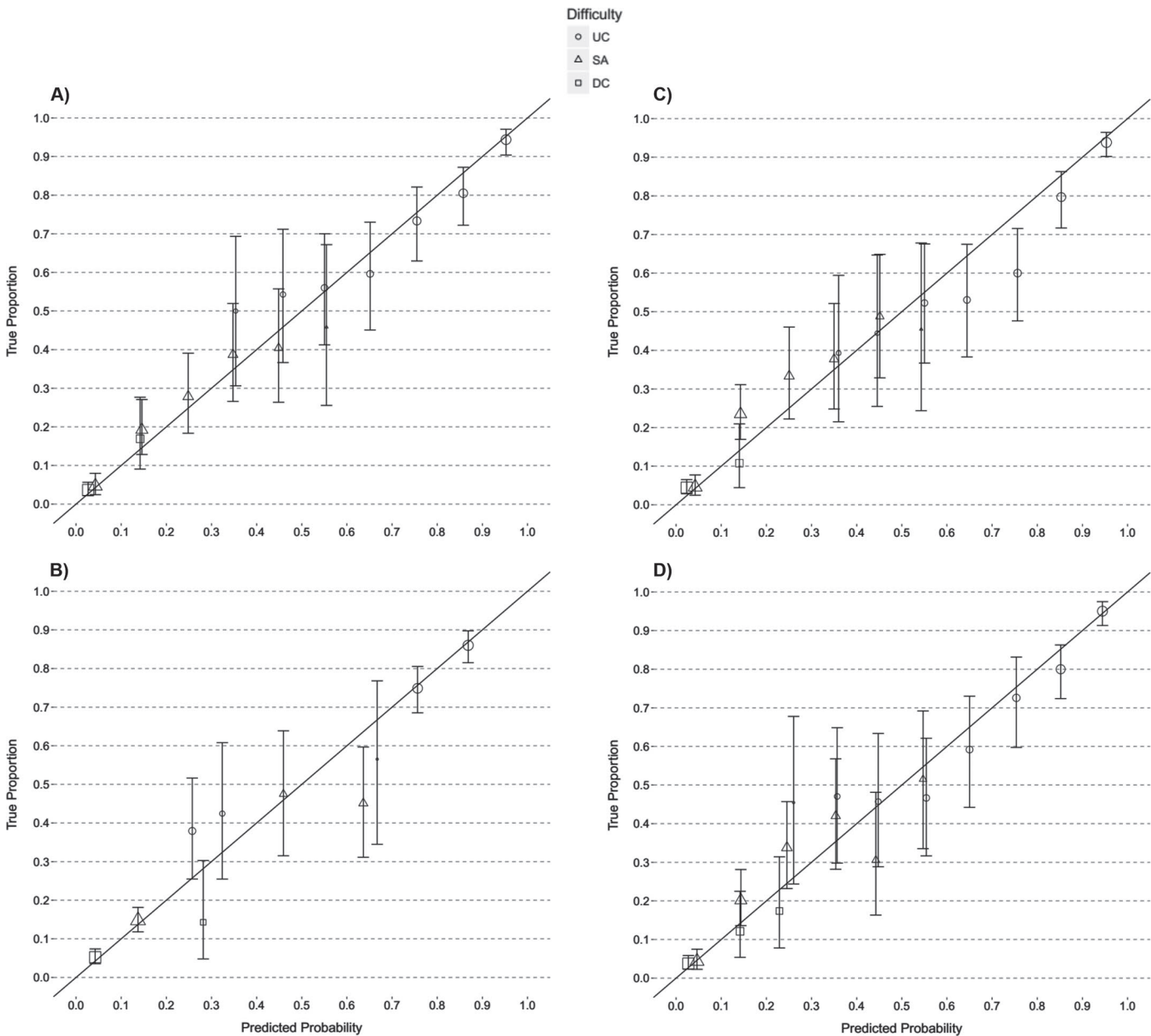
Figure 2. Calibration plots for test data predictions of each of the models. (A) Multinomial regression, (B) decision tree, (C) random forest, (D) neural network. Point size reflects the relative number of test data records in the calibration group. UC = unassisted calving; SA = slight assistance; DC = difficult calving. The error bars represent the 95% CI.

Model — DT ··· MR -- RF -· NN
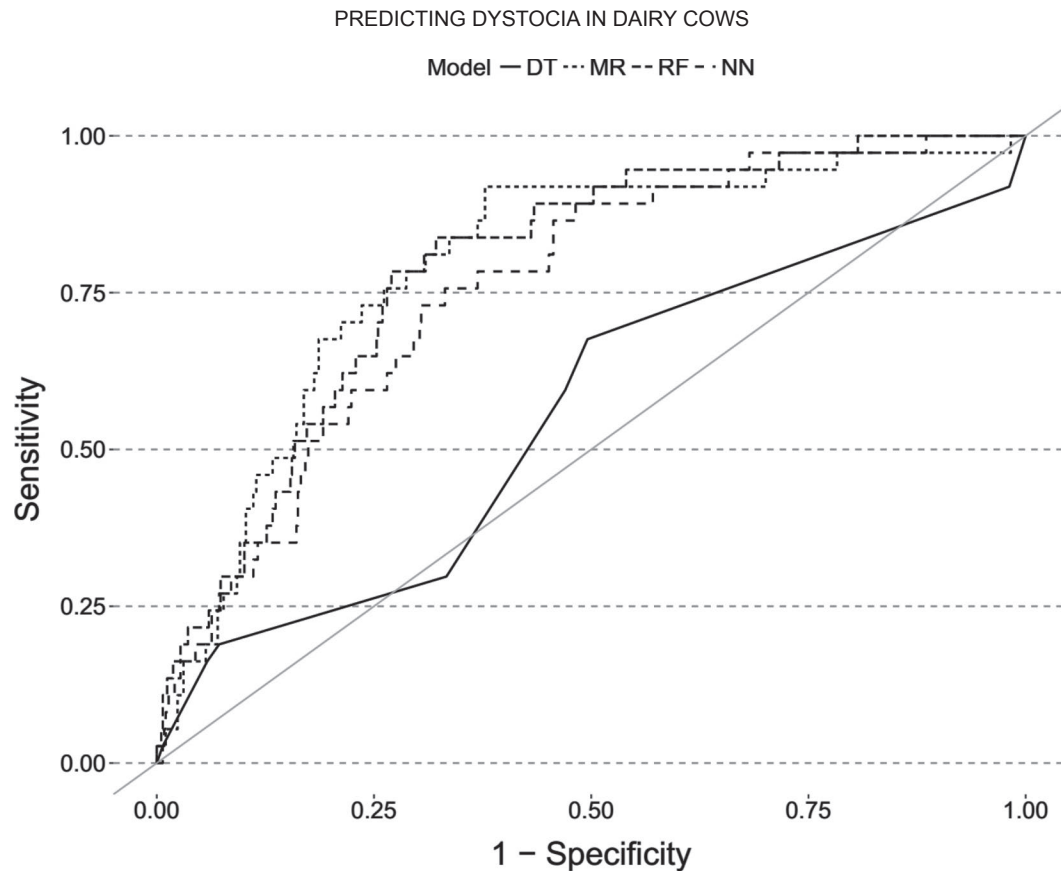
Figure 3. Receiver operating characteristic curve for difficult calving predictions in the test data set. DT = decision tree; MR = multinomial regression; RF = random forest; NN = neural network.

of our NN model performed on UC events (93%), but significantly higher than its performance on assisted calving and DC events (28% and 0%). When our models' predictions were reclassified to allow comparisons, they showed comparable performance. Classification accuracy with UC/SA versus DC was 93% (MR) and 94% (DT, RF, and NN).

The intention of our study was to create models for use in decision-support and stochastic modeling. For use as a standalone decision-support tool, excellent discriminatory performance is highly desirable, as individual animal decisions are required and probabilistic predictions are less relevant to the individual. In stochastic simulation modeling, as well as accurate discrimination, a reasonable range of predicted probabilities are desirable to allow the expression of natural variation to be simulated. For models to have utility, they should perform better than random selection and be able to accurately identify problem cows or rank cows in terms of risk. Whether it is preferable to more accurately identify cows that will have calving difficulty or those that will calve unassisted will depend on

Table 5. Model area under the receiver operating characteristic curve (AUC) and mean absolute calibration error (MACE) for test data predictions by calving difficulty level[1]

Calibration	Calving type	DT	MR	RF	NN
AUC	UC	0.69	0.79	0.80	0.79
	SA	0.67	0.76	0.77	0.77
	DC	0.55	0.79	0.79	0.76
	Model average	0.64	0.78	0.79	0.77
MACE (%)	UC	5.16	5.29	6.17	3.86
	SA	5.91	5.04	5.46	5.47
	DC	7.99	3.08	2.15	1.69
	Model average	6.35	4.47	4.59	3.67

[1]DT = decision tree; MR = multinomial regression; RF = random forest; NN = neural network; UC = unassisted calving; SA = slight assistance; DC = difficult calving.

the management strategies and facilities on the farm. For example, if only a limited number of cows can be included in a high-risk group, higher PPV for calving assistance is needed to avoid low risk cows taking the place of a high-risk cow. In contrast, if space is not a problem, a higher NPV for calving assistance is more desirable to ensure no cow is left out of the high-risk group, even at the expense of including some cows that will not require assistance. The balance in this trade-off is likely to be farm specific, and sensitivity and specificity (and therefore PPV and NPV) can be altered by changes in classification cut-offs. Given that discrimination performance can be altered to suit farm preferences, the ability of a model to correctly rank cows based on the probability of a given calving difficulty score is more important.

The F-measures were low for both SA and DC, suggesting a poor ability to identify assisted calving events. The low values were driven by the low sensitivity of the models for SA and DC outcomes (0 to 34%), indicating that many calving events were incorrectly classified as not needing any assistance. One likely reason for these results is the distribution of the outcomes, which was heavily skewed toward UC events. The DT model only predicted a narrow number of probabilities due to having only 9 terminal nodes. This makes the DT model less effective at ranking cows as it is only capable of assigning a limited range of probabilities (Figure 2).

Because this is the first application of calibration techniques to the evaluation of calving difficulty models, no comparisons to other studies can be made. The calibration methods used in the evaluation of our model indicate that both the NN and MR models can accurately predict the probability of occurrence.

For decision-support and simulation purposes, MR and DT are the most interpretable model types. As the performance of DT was poor, MR is the best proposed model. Odds ratios can be presented for each variable by difficulty level or a linear equation of the coefficients can be used to calculate the cumulative probability of calving assistance and difficulty. Our animal-level model could also be used to estimate overall rates of assistance and dystocia at the herd level, either by using average animal characteristics or by combining individual animal predictions. This could be used to facilitate benchmarking between herds.

CONCLUSIONS

In this study, we have created an accurate model of multiple levels of calving difficulty. Uniquely, we used calibration methods to select models of calving difficulty that would be suitable for use in a decision-support

tool for individual cows and in stochastic simulation modeling. We suggest that these methods should be routinely used to evaluate multiclass predictive models, in conjunction with discrimination tests. The risk factors contained in the final models used to predict the probability of calving assistance and dystocia are consistent with the known literature, giving confidence in the models' validity. The factors are based on easily recordable data, and include both genetic and phenotypic variables and interactions between them. This study also demonstrated the use of calibration evaluation techniques, which have not frequently been used in agricultural or animal health applications.

REFERENCES

Berry, D. P., J. M. Lee, K. A. Macdonald, and J. R. Roche. 2007. Body condition score and body weight effects on dystocia and stillbirths and consequent effects on postcalving performance. J. Dairy Sci. 90:4201–4211. https://doi.org/10.3168/jds.2007-0023.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen. 1984. Classification and Regression Trees. CRC Press, Boca Raton, FL.

Bureš, D., L. Bartoň, R. Zahrádková, V. Teslík, and M. Fiedlerová. 2008. Calving difficulty as related to body weights and measurements of cows and calves in a herd of Gascon breed. Czech J. Anim. Sci. 53:187–194.

Dematawewa, C. M., and P. J. Berger. 1997. Effect of dystocia on yield, fertility, and cow losses and an economic evaluation of dystocia scores for Holsteins. J. Dairy Sci. 80:754–761. https://doi.org/10.3168/jds.S0022-0302(97)75995-2.

Drew, B. 1986. Factors affecting calving rates and dystokia in Friesian dairy heifers, the results of a large scale field trial. Irish Grassl. Anim. Prod. J. 20:98–104.

Erb, H. N., R. D. Smith, P. A. Oltenacu, C. L. Guard, R. B. Hillman, P. A. Powers, M. C. Smith, and M. E. White. 1985. Path model of reproductive disorders and performance, milk fever, mastitis, milk yield, and culling in Holstein cows. J. Dairy Sci. 68:3337–3349. https://doi.org/10.3168/jds.S0022-0302(85)81244-3.

Ettema, J. F., and J. E. P. Santos. 2004. Impact of age at calving on lactation, reproduction, health, and income in first-parity Holsteins on commercial farms. J. Dairy Sci. 87:2730–2742. https://doi.org/10.3168/jds.S0022-0302(04)73400-1.

Fenlon, C., L. O'Grady, M. L. Doherty, J. Dunnion, L. Shalloo, and S. T. Butler. 2017. The creation and evaluation of a model predicting the probability of conception in seasonal-calving, pasture-based dairy cows. J. Dairy Sci. 100:5550–5563.

Fourichon, C., H. Seegers, N. Bareille, and F. Beaudeau. 1999. Effects of disease on milk production in the dairy cow: A review. Prev. Vet. Med. 41:1–35.

Fourichon, C., H. Seegers, and X. Malher. 2000. Effect of disease on reproduction in the dairy cow: A meta-analysis. Theriogenology 53:1729–1759. https://doi.org/10.1016/S0093-691X(00)00311-3.

Freeman, E. A., and G. Moisen. 2008. PresenceAbsence: An R Package for Presence Absence Analysis. J. Stat. Softw. 23:1–31. https://doi.org/10.1038/nmat2803.

Gelman, A., and Y.-S. Su. 2015. arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.8-6. Accessed Apr. 5, 2016. http://cran.rproject.org/package=arm.

Grunert, E. 1979. Clinical aspects of the nutritional status of the dam and parturition. Springer Netherlands, Dordrecht.

Hand, D. J., and R. J. Till. 2001. A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach. Learn. 45:171–186. https://doi.org/10.1023/A:1010920819831.

Horan, B., P. Dillon, P. Faverdin, L. Delaby, F. Buckley, and M. Rath. 2005. The interaction of strain of Holstein-Friesian cows and pasture-based feed systems on milk yield, body weight, and body condition score. J. Dairy Sci. 88:1231–1243. https://doi.org/10.3168/jds.S0022-0302(05)72790-9.

Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. Applied Logistic Regression. Wiley, Hoboken, NJ.

Hothorn, T., P. Buehlmann, S. Dudoit, A. Molinaro, and M. Van Der Laan. 2006. Survival ensembles. Biostatistics 7:355–373.

Huxley, J. N., and H. R. Whay. 2006. Current attitudes of cattle practitioners to pain and the use of analgesics in cattle. Vet. Rec. 159:662–668. https://doi.org/10.1136/vr.159.20.662.

Johanson, J. M., and P. J. Berger. 2003. Birth weight as a predictor of calving ease and perinatal mortality in Holstein cattle. J. Dairy Sci. 86:3745–3755. https://doi.org/10.3168/jds.S0022-0302(03)73981-2.

Keady, T. W., C. S. Mayne, D. A. Fitzpatrick, and M. A. McCoy. 2001. Effect of concentrate feed level in late gestation on subsequent milk yield, milk composition, and fertility of dairy cows. J. Dairy Sci. 84:1468–1479. https://doi.org/10.3168/jds.S0022-0302(01)70180-4.

Klassen, D. J., R. I. Cue, and J. F. Hayes. 1990. Estimation of repeatability of calving ease in Canadian Holsteins. J. Dairy Sci. 73:205–212. https://doi.org/10.3168/jds.S0022-0302(90)78665-1.

Kruip, T. A. M., and J. H. G. Den Daas. 1997. In vitro produced and cloned embryos: Effects on pregnancy, parturition and offspring. Pages 43–52 in Theriogenology. Elsevier, Amsterdam, the Netherlands.

Kuhn, M., J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt. 2016. caret: Classification and Regression Training. Accessed May 7, 2017. http://cran.r-project.org/package=caret.

LeBlanc, S. J. 2012. Interactions of metabolism, inflammation, and reproductive tract health in the postpartum period in dairy cattle. Reprod. Domest. Anim. 47:18–30. https://doi.org/10.1111/j.1439-0531.2012.02109.x.

Mee, J. F. 2008a. Newborn dairy calf management. Vet. Clin. North Am. Food Anim. Pract. 24:1–17. https://doi.org/10.1016/j.cvfa.2007.10.002.

Mee, J. F. 2008b. Prevalence and risk factors for dystocia in dairy cattle: A review. Vet. J. 176:93–101. https://doi.org/10.1016/j.tvjl.2007.12.032.

Mee, J. F., D. P. Berry, and A. R. Cromie. 2011. Risk factors for calving assistance and dystocia in pasture-based Holstein–Friesian heifers and cows in Ireland. Vet. J. 187:189–194. https://doi.org/10.1016/j.tvjl.2009.11.018.

Piwczyński, D., Z. Nogalski, and B. Sitkowska. 2013. Statistical modeling of calving ease and stillbirths in dairy cattle using the classification tree technique. Livest. Sci. 154:19–27. https://doi.org/10.1016/j.livsci.2013.02.013.

Sing, T., O. Sander, N. Beerenwinkel, and T. Lengauer. 2005. ROCR: Visualizing classifier performance in R. Bioinformatics 21:3940–3941.

Somers, J. R., J. Huxley, I. Lorenz, M. L. Doherty, and L. O'Grady. 2015. The effect of lameness before and during the breeding season on fertility in 10 pasture-based Irish dairy herds. Ir. Vet. J. 68:14. https://doi.org/10.1186/s13620-015-0043-4.

Tedeschi, L. O. 2006. Assessment of the adequacy of mathematical models. Agric. Syst. 89:225–247. https://doi.org/10.1016/j.agsy.2005.11.004.

Therneau, T., B. Atkinson, and B. Ripley. 2015. rpart: Recursive Partitioning and Regression Trees. Accessed Nov. 8, 2016. http://cran.r-project.org/package=rpart.

van Rijsbergen, C. J. 1979. Information Retrieval. 2nd ed. Butterworths, London, UK.

Venables, W. N., and B. D. Ripley. 2013. Modern Applied Statistics with S. Springer-Verlag, New York, NY.

Wu, X., V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg. 2007. Top 10 algorithms in data mining. Knowledge and Information Systems 14:1–37. https://doi.org/10.1007/s10115-007-0114-2.

Zaborski, D., and W. Grzesiak. 2011. Detection of difficult calvings in dairy cows using neural classifier. Arch. Tierzucht 54:477–489.

Zaborski, D., W. Grzesiak, K. Kotarska, I. Szatkowska, and M. Jedrzejczak. 2014a. Detection of difficult calvings in dairy cows using boosted classification trees. Indian J. Anim. Res. 48:452–458. https://doi.org/10.5958/0976-0555.2014.00010.7.

Zaborski, D., W. Grzesiak, and R. Pilarczyk. 2014b. Detection of difficult calvings in the Polish Holstein-Friesian Black-and-White heifers. J. Appl. Anim. Res. 44:42–53. https://doi.org/10.1080/09712119.2014.987293.

Zaborski, D., W. S. Proskura, and W. Grzesiak. 2016. Classification of calving difficulty scores using different types of decision trees. Acta Sci. Pol. Zootech. 15:55–70. https://doi.org/10.21005/asp.2016.15.4.05.