



Usefulness of milk mid-infrared spectroscopy for predicting lameness score in dairy cows

V. Bonfatti,^{1*} P. N. Ho,² and J. E. Pryce^{2,3}

¹Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro 35020, Italy

²Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, Victoria 3083, Australia

³School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3083, Australia

ABSTRACT

The objective of this study was to evaluate the ability of milk infrared spectra to predict cow lameness score (LMS) for use as an indicator of cow health on Australian dairy farms, or as an indicator trait for genetic evaluation purposes. The study involved 3,771 cows from 10 farms in Australia. Milk infrared spectra collected during the monthly herd testing were available in all the farms involved in the study. Lameness score was measured once in each herd, within 72 h from a test day, and merged to the closest spectra records. Lameness score was expressed on a scale from 0 to 3, where 0 is assigned to sound cows and scores 1 to 3 are assigned to cows with increased lameness severity. Partial least squares discriminant analysis was used to develop prediction models for classifying sound (score 0) and not-sound cows (i.e., cows walking unevenly, score greater than 0). Discriminant models were tested in a 10-fold random cross-validation process. Milk infrared spectra correctly classified only 57% of the cows walking unevenly and only 59% of the sound cows. When additional predictors (parity, age at calving, days in milk, and milk yield) were included in the prediction model, the model correctly classified 57% of the cows walking unevenly and 62% of the sound cows. The same model applied only to the cows in the first third of lactation correctly classified 66% of the cows walking unevenly and 57% of the sound cows. When the prediction model was used to identify lame cows (scores 2 and 3), only 49% of them were classified as such. These results are considered to be too poor to envisage a practical application of these models in the near future as on-farm tools to provide an indication of LMS. To investigate whether, at this stage, predictions of the LMS could be useful as large-scale phenotypes for animal breeding purposes, we estimated (co)variance components for actual and predicted LMS using

2,670 and 24,560 records, respectively. As the genetic correlation between actual and predicted LMS was not significantly different from zero, predictions of lameness from milk spectra and additional on-farm variables cannot be used, at this stage, as an indicator trait for actual LMS. More research is needed to find better strategies to predict lameness.

Key words: milk spectra, lameness, cow health, discriminant model

INTRODUCTION

Lameness is ranked as the third most important cause of economic loss on dairy farms, after mastitis and reproduction disorders, with a cost per animal averaging \$75 per year (Bruijnjs et al., 2010). Lame cows are more frequently affected by mastitis, metabolic disorders, and reduced fertility and milk yield (MY; Heringstad et al., 2018). Mean frequencies of lameness in dairy herds in Europe and North America range between 23 and 70% (Heringstad et al., 2018). In Australia, where most of the dairy farms are pasture based and cows may walk several kilometers and stand for hours in a crowded concrete yard while waiting to be milked, lameness is an ongoing concern. Unfortunately, farmers are able to identify only approximately 25% of the lame cows (Beggs et al., 2019).

Reducing the prevalence of lameness requires improvements in on-farm management, but selective breeding has also been explored as a long-term option (Chesnais et al., 2016; Heringstad et al., 2018). In Australia this is important, as Martin-Collado et al. (2015) indicated that lameness is one of the most important traits in future breeding objectives in a survey of Australian farmers. However, claw health traits have a low heritability and a large number of records is needed to achieve an acceptable reliability (Heringstad et al., 2018). Information related to claw health might be difficult to obtain on a large scale [e.g., computer images of claws, BCS and BW, occurrence of trimming events, housing system, diet composition, type traits; see reviews of Solano et al. (2015) and Heringstad et al.

Received September 6, 2019.

Accepted November 10, 2019.

*Corresponding author: valentina.bonfatti@unipd.it

(2018)]. For genomic selection, increasing the size of the genomic reference population is often recommended, but it is expensive and time consuming; therefore, the use of indicator traits together with actual observations could help improve the reliability (Chesnais et al., 2016).

Besides mechanical injuries or infections, lameness can be associated with physiological changes that may have a repercussion on milk composition: ketosis (Suthar et al., 2013), SARA (Oetzel, 2007), and milk fever (Sogstad et al., 2006). Hence, milk mid-infrared (MIR) spectra, which are routinely used to predict milk composition, have the potential to bear useful information for lameness detection, at least when lameness is directly a consequence of metabolic changes. Because MIR data are already routinely available (i.e., monthly) through herd testing, no extra costs would be incurred to obtain the data. Depending on the accuracy of infrared prediction models, they might be used as a tool to provide on-farm indicators of the trait at the individual level, or to define best practices and adjust feeding and management at the herd level, or as indicator traits in genetic and genomic evaluations (Bonfatti et al., 2017).

The first objective of this study was to evaluate the ability of milk MIR, together with readily available variables, including milk production, DIM, parity, and age at calving, for predicting cow lameness score (LMS). The second objective was to explore the potential of the resulting predictions as indicator traits for future genetic evaluation by estimating genetic parameters of the predicted LMS (P-LMS) and its genetic correlation with the actual score.

MATERIALS AND METHODS

Animal Data

Records of 3,771 lactating dairy cows that calved between September and October 2018 were available (data set A). Cows were from 10 commercial dairy herds located in Victoria, Tasmania, and New South Wales (Australia). Cows were predominantly Holstein-Friesian (69%), but the data set also included purebred Jersey cows (1%), Holstein-Friesian \times Jersey crossbreds (14%), and other crossbred animals (16%).

Cows on each farm were scored once, after morning milking, for LMS. Scoring was performed by 2 trained classifiers and coincided with a herd-test visit (within 72 h). Lameness was scored from 0 to 3 as per guidelines of Dairy Australia (2015). According to Beggs et al. (2019), cows with scores of 0 are assigned to walking evenly (no action is required), a score of 1 means the

cow walks unevenly (it is recommended to regularly observe the cow), while 2 and 3 are classified as lame, with 3 being severe lameness and requiring immediate attention. Similar lameness scoring systems have been shown to be robust in terms of inter- and intra-observer agreement in several studies in which researchers used each other for calibration (Barker et al., 2010; Main et al., 2010).

Available data also included parity, age at calving, DIM, MY, fat, protein, and lactose percentages, and SCS, calculated as $\log_2(\text{SCC} + 3)$. Estimated breeding values of several production and conformation traits were also available for 2,592 cows. The EBV were obtained from DataGene (Bundoora, Victoria, Australia, <https://www.datagene.com.au/>). The EBV for production traits included the balanced performance index, which is the Australian national selection index encompassing traits important for farmer profit, as described by Byrne et al. (2016), in addition to daily MY, fat yield, fat percentage, protein yield, protein percentage, and SCC. The EBV for type traits included overall type, bone quality, foot angle, rear leg rear view, rear set of leg, and live weight and are estimated using data on linear type scores as evaluated by trained operators from Holstein Australia and Jersey Australia (both located in Bundoora, Australia). Cows included in this study were part of the genomic information nucleus of females with a broad range of recorded phenotypes known as Ginfo and 70% of the cows with EBV also had their own phenotypic records for type traits. Consequently, the reliability of the estimates is expected to be relatively high.

Spectral Data

All cows were milked twice daily, which is the standard commercial practice in Australia. Milk samples collected at herd testing, which was monthly, were analyzed by TasHerd Pty Ltd. (Hadspen, Tasmania, Australia) for fat, protein, and lactose concentrations and SCC, using an infrared spectrometer (Bentley Instruments NexGen Series FTS Combi), and the corresponding spectra were stored for this study. Each recorded spectrum includes 899 data points in the 649 to 3,999 cm^{-1} region. Spectral regions characterized by low signal to noise ratio (2,998 to 3,998, 1,615 to 1,652, and 649 to 925 cm^{-1}) were removed (Hewavitharana and van Brakel, 1997).

Discriminant Models

For the discriminant analysis, records were coded binarily into sound (LMS = 0) and not-sound (i.e.,

Table 1. Predictors included in each tested model for classification of sound and lame cows

Model	Cows (no.) ¹	Parity, age at calving, DIM, milk yield	Milk composition ²	Spectrum (same day of LMS) ³	EBV for production and type traits ⁴	Spectrum (test day before LMS) ⁵
1	3,765	X				
2	3,765	X	X			
3	3,765	X		X		
4	2,592	X		X	X	
5	1,202	X		X		X

¹Number of cows after outlier elimination.

²Milk composition included the percentage of fat, protein, lactose, and SCS.

³Milk mid-infrared spectrum (515 variables) obtained on the same day of lameness scoring (LMS).

⁴The EBV for production traits included balanced performance index, Australian selection index, milk yield, fat yield, fat percentage, protein yield, protein percentage, and SCC. The EBV for type traits included overall type, bone quality, foot angle, rear leg rear view, rear set of leg, and live weight.

⁵Milk mid-infrared spectrum (515 variables) obtained on the test day before the one of LMS.

walking unevenly, LMS >0) cows. Discriminant models aiming to differentiate cows walking unevenly (positive) from sound (negative) cows were developed with partial least squares discriminant analysis (Lê Cao et al., 2011), implemented in the R package mixOmics (Rohart et al., 2017). The predictors were scaled by dividing each variable by its standard deviation using a built-in option in the package. All analyses in the present study were performed with R statistical software version 3.4.4 (R Development Core Team, 2018). The number of latent variables used in the prediction models was chosen based on the break of slope of the sum between sensitivity and specificity. The break of slope is the point where adding another latent variable does not further significantly reduce the sum between sensitivity and specificity.

Several models including different explanatory variables were tested for their capability in classifying sound and lame cows. The explanatory variables included in each model are reported in Table 1. Model 1 included features that are readily available on farm (parity, age at calving, DIM, and MY). Model 2 included the variables in model 1 and milk composition (milk fat, protein, and lactose percentages and SCC). Model 3 included the variables in model 1 and the spectrum collected within 72 h from the lameness scoring. We also investigated whether the use of EBV for production and type traits would increase the accuracy of prediction models. Hence, EBV were included in model 4, besides the predictors used in model 3. In addition to these variables, model 5 also included the full spectrum collected on the previous herd-test day (approximately 1 mo earlier). Information on EBV and the previous spectrum were not available for all animals, but only for 2,592 and 1,202 cows, respectively. Therefore, for a fair model comparison, model 3 was also applied to the same training-validation partitions used for testing

models 4 and 5. In addition, model 3 was also tested only on cows in the first third of lactation (DIM ≤120; n = 2,670; data set B).

Following the approach used by Ho et al. (2019), in order for the models to have statistically fair comparisons, a random noise matrix with dimensions of $N \times p$, where N is the number of wavenumbers in model 3 and 4 (n = 515), and p is the number of records of the validation set, was generated from a uniform distribution in the interval 0.0 to 1.0 and multiplied by a very small constant of 10^{-10} . Such a matrix was then used in model 1 and 2 to obtain several predictors comparable to the number of predictors in model 3 and 4, but containing only random noise. The matrix was also used when model 3 was applied to the same data set used for testing model 5 (n = 1,202).

In preliminary analyses, the breed was also included among the possible predictors of LMS, but it did not affect model accuracy and was removed.

The performance of each model tested was evaluated using a 10-fold random cross-validation. The data set was randomly split into 10 subsets with an equal ratio of sound cows to cows walking unevenly using the groupdata2 R package (Olsen, 2017). Then, one subset at a time was reserved for validation, whereas the remaining data were used for model training. This process was repeated 10 times until each subset was used as a validation set (i.e., 10 repeats).

Discriminant Models for Severe Lameness

Variables included in model 3 were also used to develop a model aiming at discriminating cows with LMS ≤1 from more severely lame cows (LMS ≥2). For this purpose, given the low number of positive cases, the performance of the model was evaluated by splitting the data set in 5 subsets with an equal ratio between

cows scored as $LMS \leq 1$ and cows scored as $LMS \geq 2$ and by testing the predictive ability of the model in a 5-fold random cross-validation. This model was developed on data set A and B.

Evaluation of Model Performance

The accuracy of the discriminant models was assessed for every training-validation partition by producing and calculating the area under the receiver operating characteristic curve. The optimal cut-off value for each test variable was defined as the point where the sum between sensitivity and specificity was at a maximum (i.e., equal weighing of false-positive and false-negative test results).

The statistics of the discriminant models were expressed in terms of sensitivity (the proportion of lame cows that are correctly classified as lame), specificity (the proportion of sound cows that are correctly classified as sound), and global accuracy (proportion of correct classifications on the total number of cows). A high specificity can still result in numerous false positive tests when most of the cows tested are actually sound. For this reason, the results were also expressed in terms of positive and negative predictive values (PPV and NPV, respectively). Positive predictive value is the proportion of records predicted as positive (lame) that are true positives and is calculated as

$$PPV = \frac{\text{number of true positives}}{\text{(number of true positives + number of false positives)}}.$$

Negative predictive value is the proportion of records predicted as negative (sound) that are true negatives, calculated as follows:

$$NPV = \frac{\text{number of true negatives}}{\text{(number of true negatives + number of false negatives)}}.$$

Spectral Data Preprocessing

Spectra with a global standardized Mahalanobis distance (Shenk and Westerhaus, 1995) from the population centroid greater than 4 ($n = 6$) were considered outliers and eliminated. After the outlier elimination, 3,765 spectral records and associated LMS were available for development of prediction models.

Several mathematical treatments were applied to the spectra and compared before chemometric analysis. The goal of data pretreatment is to eliminate or minimize variability unrelated to the property of interest, so that pertinent changes can be more effectively modeled

(Kohler et al., 2009). Spectra were transformed using first- and second-order derivatives, followed, or not, by extended multiplicative scatter correction. The gaps over which derivatives were calculated ranged from 1 to 5 data points and the smooth ranged from 1 to 3.

Large-Scale Predictions of Lameness Score

As most of the predictors used in this study were related to milk composition, it is reasonable to hypothesize that the resulting prediction model may benefit from the exclusion of lameness cases that are not caused by metabolic disorders, and therefore in this analysis we targeted only cows in the first 120 DIM (as in data set B). To obtain P-LMS, model 3, which was developed using data set B, was chosen and applied on a larger data set of 17,913 cows that calved from 2015 to 2018, in 35 commercial farms. The data included records of parity, age at calving, DIM, MY, and infrared spectra. The data set included 55,828 records. After editing to remove records taken after 120 DIM, eliminate potential spectra outliers as described previously, and discard cows with only 1 record, 24,560 records from 8,535 cows were retained for the genetic analysis (data set C).

Genetic Parameter Estimates for Measured and Predicted Lameness Score

Estimation of (co)variance components for LMS and P-LMS were investigated in a bivariate animal model analysis performed using the TM software (available on request from the author at andres.legarra@toulouse.inra.fr). The bivariate analysis was performed on 2,670 LMS records (data set B) and 24,560 P-LMS records (data set C). Of these, 1,872 records had both a value for LMS and P-LMS. The pedigree file was obtained from DataGene. The file was pruned to include only relatives of cows that had phenotypic records. After editing, the pedigree contained 30,797 animals from 4,068 sires and 19,332 dams.

In matrix notation, the model for LMS was

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Hq} + \mathbf{Yu} + \mathbf{Za} + \mathbf{e},$$

where \mathbf{y} is the vector of phenotypic records for LMS (0 for sound cows or 1 for cows walking unevenly); \mathbf{b} is a vector including the following effects: age at calving (from 20 to 208 mo, fitted as a covariate), and DIM (from 1 to 120, divided in 4 classes of 30 d each); \mathbf{q} is a vector of herd-test day effects, assumed to follow a multivariate normal distribution; \mathbf{u} is a vector of effects of herd of calving, year of calving, season [which

is month of calving divided into 2 groups: 1 = January to July (autumn calving) and 2 = August to December (spring calving)], all combined into a contemporary group (**HYS**), assumed to follow a multivariate normal distribution; **a** is a vector of additive genetic effects of animals assumed to follow a multivariate normal distribution; **e** is a vector of random residual effects; and **X**, **H**, **Y**, and **Z** are design matrices relating records to the corresponding effects.

For P-LMS, the model was

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{H}\mathbf{q} + \mathbf{Y}\mathbf{u} + \mathbf{S}\mathbf{pe} + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

where **y** is the vector of P-LMS phenotypic records (0 for sound cows or 1 for cows walking unevenly); **pe** is a vector of the random permanent environmental effect of the cow, to account for repeated records, assumed to be normally distributed; and **S** is a design matrix relating records to the corresponding effect. The effect of the breed was tested in preliminary analyses, but it was not significant and it was removed from the final models.

Prior distributions for effects in **b** were assumed to be uniform, whereas those in **q**, **u**, **pe**, and **a** were inverted Wishart distributions.

The threshold (or probit) models consider one or several thresholds and a liability that, over a given threshold, produces an observed phenotype. At each iteration, for each record (0 or 1), a liability is generated below or over the threshold such that the observed value is 0 or 1. For dichotomous traits, for the parameters to be identifiable, a restriction is set so that residual variance is set to 1 and threshold is set to 0. As for binary traits the residual variance is set to 1; it is not a standard inverted Wishart distribution, but a conditional inverted Wishart distribution. In multivariate analyses with several binary traits, the algorithm assumes that they are uncorrelated at the residual level.

Marginal posterior distributions of parameters of concern (i.e., covariance components, heritabilities, and correlations) were estimated by performing numerical integration through the Gibbs sampler, as implemented in the TM program. A unique Gibbs chain of 5,000,000 iterations was run. Samples were saved every 1,000 iterations, with a burn-in of 1,000,000. The posterior median was used as a point estimate of parameters of concern. Lower and upper bounds of the highest posterior density interval with 95% probability for the heritability, and correlations were obtained from the estimated marginal densities.

Estimated heritability (h^2) was defined as

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2},$$

where σ_a^2 is the additive genetic variance, and σ_e^2 corresponds to the residual variance, which is set to 1 by the algorithm.

RESULTS AND DISCUSSION

Descriptive Statistics

The frequency of the LMS in our sample and by farm is summarized in Table 2. In total, 11.9% of the cows had a LMS >0. Across farms, the proportion of cows walking unevenly varied from 3.9 to 18.7%. The incidence of lameness varies with the time of year and stage of lactation in Australia, being more of a problem in the wetter months and during early lactation (Ranjbar, 2017); thus, prevalence reported by different studies may not be directly comparable. Prevalence of lameness was reported to be 18.9% across 63 pasture-based dairy herds in New South Wales visited throughout a 12-mo period (Ranjbar et al., 2016). In Victoria, the percentage of cows that had a LMS of 2 or 3 ranged from 0 to 11.4% across 50 farms with an average of 3.8% (Beggs et al., 2019). In our sample, that percentage ranged from 0.6 to 8.8% across farms and was on average 3.8%, in line with the results of the previous study. Around 70% of the cows were in the first third of lactation. On average, cows produced 28.2 ± 8.5 kg of milk per day, containing $3.74 \pm 0.90\%$ of fat and $3.47 \pm 0.37\%$ of protein.

Accuracy of Prediction Models

Finding true positives is much more important than finding true negative cows because it is imperative to detect lame animals to treat them. A model with a high sensitivity is therefore preferred to a model with a high specificity. As reported in Appendix Table A1, in the preliminary analyses, the best performances were obtained using the spectra without the noise regions and a 2nd derivative calculated over a window of 3 points, with a smooth of 1 (no smoothing). Consequently, this was the treatment of choice, leaving 515 spectral variables for final model development. In contrast with our results, Mineur et al. (2017) observed that the 2nd derivative had a tendency to lower the sensitivity in favor of the specificity compared with the 1st derivative treatment.

Performances of predictions models are summarized in Table 3. The model including on-farm information and the spectrum (model 3) had a much higher sensitivity than the model including only on-farm information (model 1). If compared with the model that included milk composition, but not the full spectrum (model 2), model 3 provided additional information and improved

Table 2. Frequency of lameness (sound cows are defined as those with lameness score = 0; lameness score >0 indicates increased lameness severity) in the totality of data and by farm

Group	Cows (no.)	Cows walking unevenly (no.)	Lameness score frequency (%)				Proportion of cows walking unevenly (%)
			0	1	2	3	
All samples	3,771	450	88.1	8.2	3.2	0.6	11.9
Farm							
1	225	42	81.3	12.9	5.3	0.4	18.7
2	176	22	87.5	8.0	4.0	0.6	12.5
3	639	102	84.0	12.1	3.8	0.2	16.0
4	324	59	81.8	9.3	7.1	1.9	18.2
5	339	28	91.7	6.2	2.1	0.0	8.3
6	520	69	86.7	8.5	3.8	1.0	13.3
7	304	12	96.1	2.6	1.0	0.3	3.9
8	479	57	88.1	9.0	2.5	0.4	11.9
9	425	45	89.4	7.3	2.4	0.9	10.6
10	340	14	95.9	3.5	0.6	0.0	4.1

prediction performances. Neither the use of the EBV (model 4), nor the use of the spectra collected on the test day before the one of lameness scoring (model 5), improved the prediction accuracy of LMS.

The model with the highest accuracy (model 3) correctly classified 62% of the records, with 57% sensitivity and 62% specificity, when all cows were included in the analysis. When the same predictors were applied only to cows in the first 120 DIM (data set B), the model provided a slightly better result, reaching 66% sensitivity, even though with a lower specificity (56%). However, as the proportion of sound cows was much larger than the one of cows walking unevenly, the percentage of global correct classifications was lower than in data set A.

Mineur et al. (2017) obtained very similar results in terms of sensitivity and specificity in cows in any stage of lactation and cows in the first third of lactation, with a comparable prevalence of lameness, but with a larger

sample size and without including any predictors other than the spectral variables. The same study also demonstrated that using a more detailed definition of lame, by selecting the animals affected by a specific cause of lameness, improves the prediction up to a sensitivity of 85% and specificity of 91%. However, the model will then predict only that given type of lameness. As Mineur et al. (2017) did not report the PPV, it is not possible to know how many of the cows predicted as lame were truly lame. In our study, only 16% (data set A) or 14% (data set B) of the positive predictions were actually positive. A higher proportion of false positives does not necessarily result in animals that are treated for no reason (e.g., the model could be used as a screening tool, identifying cows that need further attention), but the proportion of false positives, at this stage, is too high to propose a practical application of the models.

According to the value of sensitivity, on average, in the best scenario (data set B, model 3) only 66% of the

Table 3. Performance of prediction models obtained in a 10-fold random cross-validation, for the discrimination of sound cows (negative) and cows walking unevenly (positive)¹

Model (no. of cows) ²	Optimal number of components	Sensitivity (%)	Specificity (%)	Global accuracy (%)	PPV (%)	NPV (%)	AUC (%)
1 (3,765)	2.1 ± 2.4	33.3 ± 6.0	67.9 ± 3.4	63.8 ± 2.9	12.2 ± 2.0	88.3 ± 2.0	52.3 ± 3.8
2 (3,765)	2.6 ± 2.4	34.8 ± 6.5	68.1 ± 3.6	64.1 ± 3.1	12.9 ± 2.4	88.5 ± 1.1	52.4 ± 3.4
3 (3,765)	7.6 ± 1.7	56.9 ± 7.0	62.2 ± 2.0	61.5 ± 1.9	16.9 ± 1.9	91.5 ± 1.3	62.1 ± 4.0
3 (2,670)	2.9 ± 0.9	65.7 ± 11.1	56.1 ± 4.1	57.1 ± 3.4	14.0 ± 1.8	93.8 ± 3.8	63.4 ± 5.5
3 (2,592)	7.2 ± 2.3	53.7 ± 10.8	61.7 ± 2.8	60.7 ± 2.5	18.0 ± 3.1	89.5 ± 2.1	60.0 ± 6.2
3 (1,202)	2.9 ± 1.2	36.3 ± 8.6	66.6 ± 7.1	62.0 ± 5.2	16.3 ± 2.3	85.4 ± 1.5	51.3 ± 4.3
4 (2,592)	7.0 ± 2.2	53.4 ± 12.6	61.8 ± 4.0	60.7 ± 3.4	17.9 ± 3.7	89.5 ± 2.3	60.5 ± 7.5
5 (1,202)	2.9 ± 0.9	41.7 ± 6.4	58.7 ± 8.0	56.2 ± 6.9	15.6 ± 3.4	84.8 ± 2.4	51.5 ± 5.0

¹Values correspond to the average ± SD in validation obtained for 10 training-validation partitions. Sensitivity is the proportion of positive cows classified as such; specificity is the proportion of negative cows classified as such; global accuracy is the total proportion of correctly classified records; PPV = positive predicted value (proportion of records predicted as positive that are truly positive); NPV = negative predicted value (proportion of records predicted as negative that are truly negative); AUC = area under the receiver operating characteristic curve.

²Models 4 and 5 were tested on a portion of the data (2,592 and 1,202 records, respectively) because of missing records for their predictors. To guarantee fair comparisons, model 3 was applied to the same data sets used for models 4 and 5. Model 3 was also applied only to the cows in the first third of lactation (DIM ≤120; n = 2,670).

cows walking unevenly were identified as such. Beer et al. (2016) found that models using two 3-dimensional accelerometer variables (walking speed and standing bouts) identified slightly lame cows with sensitivity and specificity both exceeding 90%. Compared with those findings, the predictive ability of the models exploiting spectral information seems to be very poor. In addition, it has to be considered that (1) there are only 2 options (sound vs. not sound), so that the random chance of classifying an animal correctly is 50%, and (2) the settings used for model validation (random splitting of the data set maintaining an equal ratio between sound and not-sound cows in each subset) represents a very conservative approach. While the random cross-validation allows for a direct comparison of our results with the literature, a herd-by-herd validation would allow to evaluate model performance across farms, representing a more realistic scenario of the model performance for consideration for a routine application, because it prevents records from the same farm to end up in both the training and validation sets and because its performance is evaluated taking into account variation in the prevalence of lameness across farms.

Accuracy of Prediction Models for Severe Lameness

For the discriminant analysis, cows were first classified as sound (LMS = 0) or walking unevenly (LMS >0), as this is the condition that requires either monitoring or action. However, a score of 1 corresponds to very mild lameness and this is very likely to be the class where most errors of classification occur. Moreover, some cows normally walk unevenly. An alternative was grouping the cows with LMS ≤1 as not lame and LMS ≥2 as lame (Table 4).

Given the limited number of scores of 2 and 3 in our data, testing this option was only possible by reducing the number of cross-validation folds. The model had a higher accuracy (global accuracy = 71%) compared with the other discriminant models. However, the high proportion of correct classifications may be attributable to the very low prevalence of lameness (only 3.7%). The

model classified the majority of records as not lame, therefore reaching a relatively high global accuracy, but it was able to identify only 44% of lame cows. The model worked slightly better when applied to cows that were in the first 120 DIM, but only 49% of the lame cows could be identified correctly. In addition, less than 5% of the cows classified as lame were actually lame.

It must be stressed that Australian farmers could correctly identify only 24% of lame cows (Beggs et al., 2019) in a study comparing expert diagnosis to farmers' diagnosis. This corresponds to a sensitivity of just 24%. So, despite having a lower global accuracy, prediction models have a higher sensitivity than farmers in classifying sound and lame cows. Farmers tend to underestimate the presence of lameness in their herds, estimating its prevalence to be roughly 30% of the actual prevalence (Beggs et al., 2019). However, even though prediction models have a better sensitivity when compared with farmers in identifying lame cows, the high proportion of false positives and low global accuracy indicate that the use of infrared spectroscopy alone is not an effective technique to predict lameness.

Infrared Predictions as a Large-Scale Phenotyping Tool for Lameness

Variance components and genetic parameters estimates obtained for LMS and P-LMS are reported in Table 5. The genetic analysis of LMS revealed that its estimated heritability was 0.022 (ranging from 0.008 to 0.059 with 95% probability), in line with a previous study on the same trait performed in Australia (Abdel-sayed et al., 2017) and other literature estimates, which ranged from 0.02 and 0.10 (Heringstad et al., 2018). The heritability of P-LMS was higher than that of the actual trait (0.167; ranging from 0.139 to 0.191 with 95% probability).

Several different sources of claw health data are now available, with the information recorded by claw trimmers showing particular promise for breeding purposes (Heringstad et al., 2018). However, due to the low heritability of these traits, many observations are needed to

Table 4. Performance of prediction models obtained in a 5-fold random cross-validation, for the discrimination of cows scored as 0 or 1 (negative) and cows scored as 2 or 3 (positive)¹

Model (no. of cows) ²	Optimal number of components	Sensitivity (%)	Specificity (%)	Global accuracy (%)	PPV (%)	NPV (%)	AUC (%)
3 (3,765)	7.0 ± 2.4	44.3 ± 11.7	72.0 ± 1.5	71.0 ± 1.3	5.8 ± 1.3	97.1 ± 0.6	62.3 ± 6.5
3 (2,670)	6.8 ± 4.3	48.7 ± 17.3	74.1 ± 9.6	73.3 ± 9.1	5.7 ± 2.0	98.0 ± 0.6	68.4 ± 8.9

¹Values correspond to the average ± SD in validation obtained for 5 training-validation partitions. Sensitivity is the proportion of positive cows classified as such; specificity is the proportion of negative cows classified as such; global accuracy is the total proportion of correctly classified records; PPV = positive predicted value (proportion of records predicted as positive that are truly positive); NPV = negative predicted value (proportion of records predicted as negative that are truly negative); AUC = area under the receiver operating characteristic curve.

²The model was applied to the entire data set (n = 3,765) or only to the cows in the first third of lactation (DIM ≤120; n = 2,670).

produce genetic evaluations with a high reliability, so the use of indicator traits that are genetically correlated with direct measures of claw health is desirable. In this scenario, P-LMS might be used as an indicator trait to improve resistance to lameness by selective breeding. This is especially true considering that even poorly predicted traits may exhibit a positive genetic correlation with the breeding goal (Bonfatti et al., 2017). However, the estimated genetic correlation between LMS and P-LMS was not different from zero (-0.052 , ranging from -0.633 to 0.563 with 95% probability). This result was obtained despite the very conservative approach used (the records with both LMS and P-LMS information were the same records used for the development of prediction model), meaning that the accuracy of prediction models is too poor to result in a positive genetic correlation between LMS and P-LMS. The lack of a positive genetic relationship between the actual trait and the prediction precludes, at this stage, the use of the infrared predictions as an indicator trait of LMS.

Obstacles Limiting Infrared Spectroscopy Model Performance

Possible reasons for the inefficacy of spectral variables to predict lameness are listed below.

- (1) Only part of the lameness events are related to metabolic changes. There are 3 main causes of lameness: (1) injury, (2) infectious diseases, and (3) metabolic disorders. Infectious and noninfectious lameness can be possibly detected by changes in milk composition. While some metabolic changes affecting the onset of lameness are known to affect milk composition and hence the spectra, it is reasonable to hypothesize

that unknown changes in milk spectra might occur as a consequence of lameness. Only lameness events related to metabolic changes can be (theoretically) predicted through infrared spectra. Models need to define precisely the target to be predicted, suggesting that the spectrum has the potential to be used only for very specific situations, as also suggested by Mineur et al. (2017). In addition, the metabolic changes causing lameness must be large enough to determine changes in milk composition.

- (2) Changes in milk composition must be relevant to be captured by spectral changes. Infrared spectra are not expected to detect small differences in milk composition, or differences in compounds present in milk in low concentrations (Eskildsen et al., 2014). For example, claw keratin formation might be correlated with the content of minerals in milk (Heringstad et al., 2018), which is not predicted accurately by infrared spectroscopy compared with the content of other compounds (Bonfatti et al., 2016).
- (3) Metabolic changes have to occur at the right moment to be detectable. Spectra collection occurs usually monthly, and in order for the model to be accurate, the interval between the occurrence of a change in milk composition and the onset of lameness should be relatively constant. Serum fatty acids show elevated proportions in milk 2 wk before ketosis (Van Haelst et al., 2008), but it is not clear how long it can take to result in lameness. When a case of lameness is a consequence of SARA, hoof lesions appear only weeks after the SARA episode (Tajik and Nazifi, 2011). In addition, the metabolic changes do not always result in lameness.

Table 5. Median and lower and upper bound of the 95% highest posterior density interval (HPD95%) of the marginal posterior density of variance components and h^2 estimated for actual lameness score (LMS) and predicted lameness score (P-LMS)

Parameter ¹	LMS		P-LMS	
	Median	HPD95%	Median	HPD95%
σ_a^2	0.023	0.008–0.062	0.201	0.162–0.237
σ_{htd}^2	0.371	0.224–0.531	0.579	0.400–0.839
σ_{hys}^2	0.139	0.015–0.399	1.228	0.612–2.037
σ_{pe}^2	—	—	0.041	0.012–0.071
h^2	0.022	0.008–0.059	0.167	0.139–0.191
r_a	—	—	-0.052	$-0.633-0.563$

¹ σ_a^2 = additive genetic variance; σ_{htd}^2 = variance of the effect of herd-test day; σ_{hys}^2 = variance of the effect of herd-year-season of calving; σ_{pe}^2 = permanent environmental variance; r_a = additive genetic correlation with LMS. The residual variance for LMS and P-LMS was set to 1 by the algorithm.

- (4) Lameness can derive from metabolic changes occurring in opposite directions. Fat can be positively or negatively associated with lameness, when lameness derives from ketosis or SARA, respectively (Plaizier et al., 2008; Loker et al., 2012). Fat mobilization (i.e., low BCS) was negatively associated with digital cushion thickness (Newsome et al., 2017) and hence reduction of the efficiency of the digital cushion as a shock absorber (Espejo et al., 2006), but high BCS was also associated with increased risk of lameness (Rouha-Mülleder et al., 2009).
- (5) Spectroscopy is an indirect technique and needs a precise reference analysis. Lameness scoring can be subject to errors, especially if cows are not completely free to show their behavior or many cows are scored simultaneously. In addition, some cows normally walk unevenly.
- (6) Not all cows that show lesions, infections, or metabolic changes show lameness. Lameness is determined through the display of a deviation from normal behavior (i.e., it is primarily determined by the amount of pain a cow experiences, not directly from the severity of a lesion, infection, or metabolic disorder).
- (7) Lameness events have a relatively low prevalence. The unbalanced data set in model training is generally a source for poor model performance, which also applies to other traits (e.g., BHB; Bonfatti et al., 2019).
- (8) Lameness events are greatly determined by on-farm management. The different sources of variation that exist in the field (e.g., farm, housing system, flooring type, and so on) are not taken into account when using infrared spectra as the sole predictor of LMS.

Considering these limitations, prediction of specific claw disorders may work better than a prediction of the more general LMS. Moreover, a better option could be to also use data from activity sensors, which have been found to be good indicators of lameness (Beer et al., 2016), to include more information on farming/housing conditions, and to expand the database to include a higher number of severe lameness cases.

CONCLUSIONS

The accuracy of prediction models for LMS exploiting milk infrared spectral variables and additional readily available on-farm predictors was considered too poor to envisage a practical application of these models in the near future as on-farm tools to provide an indica-

tion of cow lameness status. In addition, the genetic correlation between the actual LMS and the P-LMS was not significantly different from zero. Therefore, at this stage, predictions of the LMS cannot be used as large-scale phenotypes for animal breeding purposes, as an indicator trait for actual lameness. The results showed that the large-scale prediction of lameness in cows still needs additional research. The incorporation in the prediction models of data from activity sensors (e.g., accelerometers) and other farming/housing information, or the prediction of the occurrence of specific claw disorders as an alternative to the LMS, should be investigated.

ACKNOWLEDGMENTS




The authors thank DairyBio, jointly funded by Dairy Australia (Melbourne, Australia), The Gardiner Foundation (Melbourne, Australia), and Agriculture Victoria (Melbourne, Australia), for funding this project. The staff of DataGene (Bundoora, Victoria, Australia) are gratefully acknowledged for organizing and collecting LMS and extracting the data used in this study. The authors also thank colleagues at Agriculture Victoria who helped with this study in a variety of ways. Finally, we acknowledge the herd-test organizations and the farmers involved in this study. The authors have not stated any conflicts of interest.

REFERENCES

- Abdelsayed, M., M. Haile-Mariam, and J. E. Pryce. 2017. Genetic parameters for health traits using data collected from genomic information nucleus herds. *J. Dairy Sci.* 100:9643–9655. <https://doi.org/10.3168/jds.2017-12960>.
- Barker, Z. E., K. A. Leach, H. R. Whay, N. J. Bell, and D. C. Main. 2010. Assessment of lameness prevalence and associated risk factors in dairy herds in England and Wales. *J. Dairy Sci.* 93:932–941. <https://doi.org/10.3168/jds.2009-2309>.
- Beer, G., M. Alsaad, A. Starke, G. Schuepbach-Regula, H. Müller, P. Kohler, and A. Steiner. 2016. Use of extended characteristics of locomotion and feeding behavior for automated identification of lame dairy cows. *PLoS One* 11:e0155796. <https://doi.org/10.1371/journal.pone.0155796>.
- Beggs, D. S., E. C. Jongman, P. E. Hemsworth, and A. D. Fisher. 2019. Lame cows on Australian dairy farms: A comparison of farmer-identified lameness and formal lameness scoring, and the position of lame cows within the milking order. *J. Dairy Sci.* 102:1522–1529. <https://doi.org/10.3168/jds.2018-14847>.
- Bonfatti, V., L. Degano, A. Menegoz, and P. Carnier. 2016. Short communication: Mid-infrared spectroscopy prediction of fine milk composition and technological properties in Italian Simmental. *J. Dairy Sci.* 99:8216–8221. <https://doi.org/10.3168/jds.2016-10953>.
- Bonfatti, V., S.-A. Turner, B. Kuhn-Sherlock, T. D. W. Luke, P. N. Ho, C. V. C. Phyn, and J. E. Pryce. 2019. Prediction of blood β -hydroxybutyrate content and occurrence of hyperketonemia in early-lactation, pasture-grazed dairy cows using milk infrared spectra. *J. Dairy Sci.* 102:6466–6476. <https://doi.org/10.3168/jds.2018-15988>.
- Bonfatti, V., D. Vicario, A. Lugo, and P. Carnier. 2017. Genetic parameters of measures and population-wide infrared predictions of

- 92 traits describing the fine composition and technological properties of milk in Italian Simmental cattle. *J. Dairy Sci.* 100:5526–5540. <https://doi.org/10.3168/jds.2016.11667>.
- Bruijnjs, M. R. N., H. Hogeveen, and E. N. Stassen. 2010. Assessing economic consequences of foot disorders in dairy cattle using a dynamic stochastic simulation model. *J. Dairy Sci.* 93:2419–2432. <https://doi.org/10.3168/jds.2009.2721>.
- Byrne, T. J., B. F. S. Santos, P. R. Amer, D. Martin-Collado, J. E. Pryce, and M. Axford. 2016. New breeding objectives and selection indices for the Australian dairy industry. *J. Dairy Sci.* 99:8146–8167. <https://doi.org/10.3168/jds.2015.10747>.
- Chesnais, J. P., T. A. Cooper, G. R. Wiggans, M. Sargolzaei, J. E. Pryce, and F. Miglior. 2016. Using genomics to enhance selection of novel traits in North American dairy cattle. *J. Dairy Sci.* 99:2413–2427. <https://doi.org/10.3168/jds.2015.9970>.
- Dairy Australia. 2015. Healthy Hooves: Investigation Pack. 1st ed. Dairy Australia Ltd., Southbank, Victoria, Australia.
- Eskildsen, C. E., M. A. Rasmussen, S. B. Engelsen, L. B. Larsen, N. A. Poulsen, and T. Skov. 2014. Quantification of individual fatty acids in bovine milk by infrared spectroscopy and chemometrics: Understanding predictions of highly collinear reference variables. *J. Dairy Sci.* 97:7940–7951. <https://doi.org/10.3168/jds.2014.8337>.
- Espejo, L. A., M. I. Endres, and J. Salfer. 2006. Prevalence of lameness in high-producing Holstein cows housed in freestall barns in Minnesota. *J. Dairy Sci.* 89:3052–3058. [https://doi.org/10.3168/jds.S0022-0302\(06\)72579-6](https://doi.org/10.3168/jds.S0022-0302(06)72579-6).
- Heringstad, B., C. Egger-Danner, N. Charfeddine, J. E. Pryce, K. F. Stock, J. Kofler, A. M. Sogstad, M. Holzhauser, A. Fiedler, K. Müller, P. Nielsen, G. Thomas, N. Gengler, G. de Jong, C. Ødegård, F. Malchiodi, F. Miglior, M. Alsaad, and J. B. Cole. 2018. Invited review: Genetics and claw health: Opportunities to enhance claw health by genetic selection. *J. Dairy Sci.* 101:4801–4821. <https://doi.org/10.3168/jds.2017.13531>.
- Hewavitharana, A. K., and B. van Brakel. 1997. Fourier transform infrared spectrometric method for rapid determination of casein in raw milk. *Analyst (Lond.)* 122:701–704. <https://doi.org/10.1039/a700953d>.
- Ho, P. N., V. Bonfatti, T. D. W. Luke, and J. E. Pryce. 2019. Classifying fertility of dairy cows using milk mid-infrared spectroscopy. *J. Dairy Sci.* 102:10460–10470. <https://doi.org/10.3168/jds.2019.16412>.
- Kohler, A., M. Zimonja, V. Segtnan, and H. Martens. 2009. 2.09 - Standard normal variate, multiplicative signal correction and extended multiplicative signal correction preprocessing in biospectroscopy. Pages 139–162 in *Comprehensive Chemometrics*. S. D. Brown, R. Tauler, and B. Walczak, ed. Elsevier, Oxford, UK.
- Lê Cao, K.-A., S. Boitard, and P. Besse. 2011. Sparse PLS discriminant analysis: Biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics* 12:253–286. <https://doi.org/10.1186/1471-2105-12-253>.
- Loker, S., C. Bastin, F. Miglior, A. Sewalem, L. R. Schaeffer, J. Jamrozik, A. Ali, and V. Osborne. 2012. Genetic and environmental relationships between body condition score and milk production traits in Canadian Holsteins. *J. Dairy Sci.* 95:410–419. <https://doi.org/10.3168/jds.2011.4497>.
- Main, D. C. J., Z. E. Barker, K. A. Leach, N. J. Bell, H. R. Why, and W. J. Browne. 2010. Sampling strategies for monitoring lameness in dairy cattle. *J. Dairy Sci.* 93:1970–1978. <https://doi.org/10.3168/jds.2009.2500>.
- Martin-Collado, D., T. J. Byrne, P. R. Amer, B. F. S. Santos, M. Axford, and J. E. Pryce. 2015. Analyzing the heterogeneity of farmers' preferences for improvements in dairy cow traits using farmer typologies. *J. Dairy Sci.* 98:4148–4161. <https://doi.org/10.3168/jds.2014.9194>.
- Mineur, A., A. Köck, C. Grelet, N. Gengler, C. Egger-Danner, and J. Sölkner. 2017. First results in the use of milk mid-infrared spectra in the detection of lameness in Austrian dairy cows. *ACS Agric. Conspec. Sci.* 82:163–166. <https://hrca.srce.hr/191814>.
- Newsome, R. F., M. J. Green, N. J. Bell, N. J. Bollard, C. S. Mason, H. R. Why, and J. N. Huxley. 2017. A prospective cohort study of digital cushion and corium thickness. Part 1: Associations with body condition, lesion incidence, and proximity to calving. *J. Dairy Sci.* 100:4745–4758. <https://doi.org/10.3168/jds.2016.12012>.
- Oetzel, G. R. 2007. Subacute ruminal acidosis in dairy herds: Physiology, pathophysiology, milk fat responses and nutritional management. Pages 89–119 in *Am. Assoc. Bovine Pract. 40th Annual Conf.*, Vancouver, BC, Canada. University of Wisconsin, Madison.
- Olsen, R. L. 2017. Subsetting methods for balanced cross-validation, time series windowing, and general grouping and splitting of data. Accessed 09–01–2019. <http://www.et.bs.edu/es/cran/web/packages/groupdata2/groupdata2.pdf>.
- Plaizier, J. C., D. O. Krause, G. N. Gozho, and B. W. McBride. 2008. Subacute ruminal acidosis in dairy cows: The physiological causes, incidence and consequences. *Vet. J.* 176:21–31. <https://doi.org/10.1016/j.tvjl.2007.12.016>.
- R Development Core Team. 2018. The GNU Project. The R Project for Statistical Computing. Accessed Nov. 4, 2018. <https://www.r-project.org/>.
- Ranjbar, S. 2017. Lameness in Pasture-Based Dairy Farms in NSW, Australia. PhD Thesis. Faculty of Veterinary Science, University of Sydney, Sydney, Australia.
- Ranjbar, S., A. R. Rabiee, A. Gunn, and J. K. House. 2016. Identifying risk factors associated with lameness in pasture-based dairy herds. *J. Dairy Sci.* 99:7495–7505. <https://doi.org/10.3168/jds.2016.11142>.
- Rohart, F., B. Gautier, A. Singh, and K.-A. Lê Cao. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13:e1005752. <https://doi.org/10.1371/journal.pcbi.1005752>.
- Rouha-Mülleler, C., C. Iben, E. Wagner, G. Laaha, J. Troxler, and S. Waiblinger. 2009. Relative importance of factors influencing the prevalence of lameness in Austrian cubicle loose-housed dairy cows. *Prev. Vet. Med.* 92:123–133. <https://doi.org/10.1016/j.prevetmed.2009.07.008>.
- Shenk, J. S., and M. O. Westerhaus. 1995. Forage analysis by near infrared spectroscopy. Pages 111–120 in *Forages*. Vol. II. The Science of Grassland Agriculture. 5th ed. R. F. Barnes, D. A. Miller, and C. J. Nelson, ed. Iowa State University Press, Ames.
- Sogstad, Å. M., O. Østerås, and T. Fjeldaas. 2006. Bovine claw and limb disorders related to reproductive performance and production diseases. *J. Dairy Sci.* 89:2519–2528. [https://doi.org/10.3168/jds.S0022-0302\(06\)72327-X](https://doi.org/10.3168/jds.S0022-0302(06)72327-X).
- Solano, L., H. W. Barkema, E. A. Pajor, S. Mason, S. J. LeBlanc, J. C. Zaffino Heyerhoff, C. G. R. Nash, D. B. Haley, E. Vasseur, D. Pellerin, J. Rushen, A. M. de Passillé, and K. Orsel. 2015. Prevalence of lameness and associated risk factors in Canadian Holstein-Friesian cows housed in freestall barns. *J. Dairy Sci.* 98:6978–6991. <https://doi.org/10.3168/jds.2015.9652>.
- Suthar, V. S., J. Canelas-Raposo, A. Deniz, and W. Heuwieser. 2013. Prevalence of subclinical ketosis and relationships with postpartum diseases in European dairy cows. *J. Dairy Sci.* 96:2925–2938. <https://doi.org/10.3168/jds.2012.6035>.
- Tajik, J., and S. Nazifi. 2011. Diagnosis of subacute ruminal acidosis: A review. *Asian J. Anim. Sci.* 5:80–90. <https://doi.org/10.3923/ajas.2011.80.90>.
- Van Haelst, Y. N. T., A. Beeckman, A. T. M. Van Knegsel, and V. Fievez. 2008. Short communication: Elevated concentrations of oleic acid and long-chain fatty acids in milk fat of multiparous subclinical ketotic cows. *J. Dairy Sci.* 91:4683–4686. <https://doi.org/10.3168/jds.2008.1375>.

ORCID

- V. Bonfatti  <https://orcid.org/0000-0003-3970-5764>
 P. N. Ho  <https://orcid.org/0000-0001-9481-9612>
 J. E. Pryce  <https://orcid.org/0000-0002-1397-1282>

APPENDIX

Table A1. Validation accuracy (mean \pm SD) of the partial least squares discriminant analysis models for classifying cow lameness score when spectral data were transformed according to different mathematical treatments¹

Mathematical treatment ²	10-fold random cross-validation		
	Sensitivity ³	Specificity ⁴	Global accuracy ⁵
Raw spectra	52.9 \pm 7.2	61.3 \pm 2.5	60.3 \pm 2.0
1st derivative, 321	53.3 \pm 6.2	62.8 \pm 3.0	61.6 \pm 2.5
2nd derivative, 111	54.2 \pm 5.8	58.8 \pm 3.4	58.3 \pm 2.8
2nd derivative, 532	55.3 \pm 6.3	59.7 \pm 1.9	59.2 \pm 1.9
2nd derivative, 321	57.4 \pm 6.2	62.0 \pm 2.1	61.5 \pm 2.1
2nd derivative, 321 + EMSC	57.4 \pm 5.2	61.1 \pm 2.1	60.6 \pm 1.8
3rd derivative, 321	55.8 \pm 8.8	60.5 \pm 4.4	59.9 \pm 3.1

¹All models included milk yield, age at calving, parity, days in lactation, and the infrared spectrum taken within 72 h of the scoring for lameness.

²The sequence of numbers represents, respectively, the order of the derivative, the gap over which the derivative was calculated, and the data points used for smoothing; EMSC = extended multiplicative scatter correction.

³Sensitivity = proportion of positive cows that were correctly classified.

⁴Specificity = proportion of negative cows that were correctly classified.

⁵Global accuracy is the total proportion of correctly classified records.