Prediction of likelihood of conception in dairy cows using milk mid-infrared spectra collected before the first insemination and machine learning algorithms

W. Lou,[1,2,3] **V. Bonfatti,**[4]* **H. Bovenhuis,**[2] **R. Shi,**[1,2,3] **A. van der Linden,**[3] **H. A. Mulder,**[2] **L. Liu,**[5] **Y. Wang,**[1]* **and B. Ducro**[2]

[1]State Key Laboratory of Animal Biotech Breeding, National Engineering Laboratory of Animal Breeding, Key Laboratory of Animal Genetics, Breeding and Reproduction, College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China
[2]Wageningen University & Research, Animal Breeding and Genomics, 6700 AH Wageningen, the Netherlands
[3]Wageningen University & Research, Animal Production Systems, 6700 AH Wageningen, the Netherlands
[4]Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro, 35020, Italy
[5]Beijing Dairy Cattle Center, Beijing, 100192, China

ABSTRACT

Accurate and ex-ante prediction of cows' likelihood of conception (LC) based on milk composition information could improve reproduction management on dairy farms. Milk composition is already routinely measured by mid-infrared (MIR) spectra, which are known to change with advancing stages of pregnancy. For lactating cows, MIR spectra may also be used for predicting the LC. Our objectives were to classify the LC at first insemination using milk MIR spectra data collected from calving to first insemination and to identify the spectral regions that contribute the most to the prediction of LC at first insemination. After quality control, 4,866 MIR spectra, milk production, and reproduction records from 3,451 Holstein cows were used. The classification accuracy and area under the curve (AUC) of 6 models comprising different predictors and 3 machine learning methods were estimated and compared. The results showed that partial least square discriminant analysis (PLS-DA) and random forest had higher prediction accuracies than logistic regression. The classification accuracy of good and poor LC cows and AUC in herd-by-herd validation of the best model were 76.35% ± 10.60% and 0.77 ± 0.11, respectively. All wavenumbers with values of variable importance in the projection higher than 1.00 in PLS-DA belonged to 3 spectral regions, namely from 1,003 to 1,189, 1,794 to 2,260, and 2,300 to 2,660 cm^{-1}. In conclusion, the model can predict LC in dairy cows from a high productive TMR system before insemination with a relatively good accuracy, allowing farmers to intervene in advance or adjust the insemination schedule for cows with a poor predicted LC.
Key words: likelihood of conception, milk mid-infrared spectra, discriminant analyses, cow reproduction

INTRODUCTION

Reproduction is an essential element of profitable dairy farming (Cabrera, 2014; Kaniyamattam et al., 2016). Cows that do not get pregnant after repeated inseminations are often culled because their future profit is expected to be substantially lower than that of replacement heifers due to the prolonged calving interval and lower milk revenue. As one of the most critical traits in reproductive efficiency, the likelihood of conception (**LC**) is the combined consequence of fertilization, early embryonic, late embryonic, and fetal development (Grimard et al., 2006). Each of these steps in establishing conception and pregnancy may be affected by the environment and farm management. A previous study estimated that the economic loss of each additional day open is about $5.12 to $6.00 in high-yielding dairy cows in confined systems with TMR (Cabrera, 2014). Prior knowledge of the LC to a given insemination would enable farmers to optimize breeding decisions. If a cow's LC is low, farmers may decide to skip the insemination and wait for a subsequent estrus event, treat the cow, opt for culling at a later stage of lactation, or inseminate multiple semen doses and use high-fertility semen to potentially improve the calving rate (**CR**). Thus, ex-ante identification of cows with low LC might contribute to implementing more effective reproductive strategies.

In addition to genetic factors, many nongenetic factors are known to be associated with cow LC, such as parity (Berry et al., 2011), uterine diseases (Gröhn and Rajala-Schultz, 2000), bull semen quality (DeJarnette et al.,

2004), milk protein concentration (Morton et al., 2017), milk yield (Rearte et al., 2018), and milk progesterone concentration (Blavy et al., 2018). Accordingly, several studies have attempted to use information on these nongenetic factors to predict the LC of dairy cows and obtained an area under the curve (**AUC**) ranging from 0.63 to 0.91 for classification accuracy in dairy cows (Grzesiak et al., 2010; Shahinfar et al., 2014; Blavy et al., 2018). However, some of the input variables of the LC prediction models presented in these studies, such as progesterone, may not be easily available on-farm. A lack of data thus limits the large-scale application of these models.

Mid-infrared (**MIR**) spectroscopy is a high-throughput, cost-effective tool, that is available worldwide as it provides quantitative information on the major milk constituents (i.e., fat, protein, and lactose) as part of routine milk testing (De Marchi et al., 2014). The MIR spectra have been used to develop prediction equations for methane emissions (Vanrobays et al., 2016), pregnancy status (Toledo-Alvarado et al., 2018a; Brand et al., 2021), lameness (Bonfatti et al., 2020), and many other novel economic traits in dairy cattle. Prediction equations are based on evidence that milk MIR spectra can contain information on the individual physiological status and that the physiological status influences milk composition. (Benedet et al., 2019; Ho et al., 2021; Franzoi et al., 2023). As changes in milk composition related to changes in sex hormone concentrations occur before insemination (Toledo-Alvarado et al., 2018b), milk MIR spectra were used in a few studies to predict LC in dairy cows before insemination. These studies reported AUC values between 0.487 and 0.675 (Hempstalk et al., 2015), and classification accuracies between 0.66 and 0.75 from studies on Holstein-Friesian and purebred Jersey populations in pasture-based systems in Ireland and Australia, respectively (Ho et al., 2019; Ho and Pryce, 2020). Also, the MIR-predicted LC had a greater heritability than true LC phenotypes in an Australia pasture system (van den Berg et al., 2021). These results highlighted the potential of milk MIR spectra as a predictor of LC, and of MIR-predicted reproductive traits as indicator traits in genetic evaluations of fertility. Quality of MIR-predicted LC has mainly been studied in pasture systems, and it would be of interest to analyze the performance in more intensive systems. To our knowledge, no studies have reported results based on MIR spectra collected from calving to first insemination in high productive TMR systems. In addition, the ability of different regression methods to predict LC has not yet been investigated. An in-depth knowledge of machine learning tool performance might increase the prediction accuracy of LC with MIR spectra, facilitate the future application of this strategy, and reduce additional costs in farms. Therefore, this study aimed to (1) assess the ability of 3 machine learning methods to classify cows based on their LC at first insemination using MIR spectra and on-farm predictors in highly productive Chinese Holstein cows, and (2) identify the most relevant wavenumbers in the MIR spectrum used for the classification of LC.

MATERIALS AND METHODS

Datasets

All data were collected from December 2017 to April 2021 from 18 commercial freestall farms around Beijing area (Hebei Province, China), where dairy cows were fed a TMR. The average (±SD) test-day milk yield (**MY**) per cow was 35.87 ± 10.57 kg/d. Monthly test-day spectral records were collected using 3 Fourier Transform Spectrometers (Bentley Instruments, Chaska, MN) at the Beijing Dairy Cattle Centre (Beijing, China). Each record contained 899 wavenumbers, where each wavenumber represents the absorption of infrared light through the milk sample at a particular wavelength in the range from 649.03 to 3,999.59 cm^{-1}. Test-day records for individual cows included DIM, parity, MY, milk fat percentage (**FP**), protein percentage (**PP**), lactose percentage (**LP**), milk solids, SCC, BHB in blood, MUN, milk freezing point depression (**FPD**), and contents of the main fatty acids or groups of fatty acids, namely MUFA, PUFA, SFA, UFA, C16:0, C18:0, and C18:1n-6. All milk components were predicted from milk MIR spectra via built-in modules of the Fourier transform spectrometers. Furthermore, reproductive records (insemination events, pregnancy diagnosis events, and calving events) were collected from 123,806 cows on the same 18 farms from 2010 to 2021. In total, MIR spectra, test-day, and reproduction records were available for 12,430 Holstein cows. The average (±SD) number of cows and records in each farm was 690 ± 685 and 5,024 ± 6,448, respectively.

Definition of LC and Data Editing

Following the methodology of Ho et al. (2019) and Ho and Pryce (2020), 3 classes of LC at first insemination were distinguished: good (if the cow conceived at first insemination), average (if the cow conceived following 2 or more inseminations), and poor (for cows without a subsequent conception event recorded). The conception status to a given insemination was determined based on (1) whether or not a subsequent insemination existed for that lactation, (2) the pregnancy diagnosis outcome by B-ultrasound, and (3) the subsequent calving date. The last insemination with a positive pregnancy test result or with a subsequent calving record was considered a conception. To exclude outliers, only the reproductive records corresponding to a gestation length ranging from

268 to 296 d and an interval from calving to first insemination (**ICF**) ranging from 20 to 230 d were kept in the analyses. Only data on complete lactations were kept avoiding the risk that missed pregnancy diagnosis and next calving recording would result in misclassification of a conception event as a nonconception event. Then, only the MIR spectra collected between calving and first insemination in cows up to 4 parities were selected for the subsequent analyses. To further characterize the LC groups, we compared the means of DIM, ICF, MY, and other milk compositions between the groups and checked for significant differences using the GLM procedure in SAS (v 9.21). The assumption of normality was not met for the variables DIM, ICF and SCC. We therefore performed an ANOVA on the log-transformed values of DIM and ICF and on the SCS instead for evaluating significant differences. The test results remained the same on the transformed and original scale and for ease of interpretation we presented the results on the original scale.

Absorbance values in MIR spectra were smoothed by the most effective method of Savitzky-Golay second-order derivative method (with a window size of 7; Savitzky and Golay, 1964) using the 'prospectr' package (Mullen and Stokkum, 2007) in R software (version 4.0.3) after comparison of different preprocessing algorithms. This procedure is commonly used to remove noise and perturbations from water and other molecules in milk (Grelet et al., 2021). Based on previous studies and known water-related regions, 538 wavenumbers (928–1,596 cm^{-1} and 1,693–3,025 cm^{-1}) out of the 899 were kept for further analyses (Grelet et al., 2016). Following the review of Grelet et al. (2021), the outlier removal was performed after preprocessing, using the standardized Mahalanobis distance (often known as global H distance, **GH**) combined with principal component (**PC**) analysis (Grelet et al., 2021) as follows:

$$GH = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^{\mathbf{T}}\mathbf{S}^{-1}(\mathbf{x} - \boldsymbol{\mu})}\Big/nPC,$$

where *GH* is the standardized Mahalanobis distance; **x** is the matrix of PC scores; **μ** is the mean of PC scores; **T** is the transpose in matrix operations; **S** is the (co-)variance matrix between PC scores; and *nPC* is the number of PC considered for the calculation, which explained 99% of the total variance. To prevent loss of many spectra and to retain useful information, 95% of the data (4,866 records from 3,451 cows) were kept for further analysis after GH outlier elimination. Milk spectra collected from the same animal at different times were assumed to be unique. Thus, multiple milk sample records per cow (i.e., 1.4 on average) were considered in this study. The good, average, and poor LC groups included, respectively, 1,021, 2,100, and 1,745 records.

Models, Machine Learning Algorithms, and Evaluation of Model Performance

Six models comprising different predictors were tested for their capability to classify cows with different LC (Table 1). Model 1 included basic cow registration information, which is easily available on farms. These included DIM at test day, parity, calving month, and ICF. In addition to the predictors used in model 1, model 2 included traditional milk composition traits (FP, PP, LP, and milk solids), MY, and SCC. Model 3 included the traditional milk composition and additional traits (BHB, MUN, FPD, MUFA, PUFA, SFA, UFA, C16:0, C18:0, and C18:1n-6). In addition to these predictors of model 3, model 4 also included basic cow registration information. Model 5 included the MIR spectra only, whereas model 6 included, in addition to the MIR spectra, the basic registration of model 1 as well as MY and SCC.

Logistic regression (**LR**), partial least squares discriminant analysis (**PLS-DA**), and random forest (**RF**) methodologies have been employed in the construction of discriminant models for LC classification because of their extensive application in classification tasks and notable efficacy in predictive accuracy (Hempstalk et al., 2015). Logistic regression was performed using the *glm* function in the R package "Stats." The maximum number of iterations in LR was set at 200. The "Caret" package was used to perform PLS-DA (Kuhn, 2008), testing 1 up to 538 latent factors. Random forest classification was performed using the "randomForest" package (Breiman, 2001). Model tuning was performed by testing with 100 up to 300 trees and 5 up to 40 features. The optimal parameters for each of the 3 machine learning methods were chosen to maximize the total classification accuracy and the AUC in cross-validation.

The performance of the 6 models and 3 machine learning methods was evaluated by random and herd-by-herd cross-validations. However, the group with average LC was dominant, and there is particular interest in predicting cows with poor LC, rather than an overall prediction across all groups. Hence, the average LC group was excluded from the training process. This simple synthetic minority over-sampling technique (**SMOTE**) was adopted to increase model robustness toward the detection of extreme observations. Hence, only cows from the good and poor LC groups were used in model training.

In the random cross-validation, the data from 17 out of 18 farms was randomly divided into 10 subsets of equal size, and one subset at a time was holdout for internal testing (the testing set included good, average, and poor LC cows). All the other subsets were selected to train the model (the training set included good and poor LC cows) and this process was repeated 10 times until all the subsets were used as testing set. The records from the

remaining one farm were used for herd-by-herd validation (including good, average, and poor LC cows) based on the model from internal validation. This process was repeated 18 times until each farm was used as a testing set in the herd-by-herd validation, and the predicted values were used to estimate the model classification performance.

The above process was used to verify the performance of 2-classification models (good vs. poor LC). Total classification accuracy, AUC, specificity, and sensitivity were used to measure model performance (Equations 1, 2, 3).

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN), \quad [1]$$

$$\text{Sensitivity} = TP/(TP + FN), \quad [2]$$

$$\text{Specificity} = TN/(TN + FP), \quad [3]$$

where TP, TN, FP, and FN were the total number of true positives, true negatives, false positives, and false negatives in model predictions, respectively. Here a "positive" was defined as a cow with good LC and a "negative" was a cow with poor LC. The average values for each of the indicators across subsets (in the random cross-validation) and across farms (in the herd-by-herd validation) were calculated and compared. The middle group was not considered in the metrics calculation. The most relevant predictors of LC at first insemination were identified by extracting the variable importance in projection (**VIP**) values from the model exhibiting the best performance (Mehmood et al., 2012).

Physical factors including instrument wear and variations in milk testing environment lead to differences in the response among spectrometers over time (Grelet et al., 2015). Such differences might be confounded with the sample batch (e.g., when samples from different farms are collected on different days) and lead to an overestimation of MIR prediction accuracy, as demonstrated by Wang and Bovenhuis (2019). Hence, a prediction of LC obtained using only the water-related MIR regions and LR, PLS-DA, and RF methods was employed as a negative control. As water-absorbance regions contain noise, their classification accuracy should not be different from a random classification (Wang and Bovenhuis, 2019). The chi-squared test was used to check whether the accuracies were significantly different from 0.5. All the above calculations were performed with the R (v 4.0.3; R Core Team) and SAS (v 9.2; SAS Institute Inc.) software.

RESULTS

Descriptive Statistics

Figure 1 presents the distribution of the CR at first insemination in the sampled farms before and after the

Table 1. Variables used as predictors in models to classify cows with good and poor likelihood of conception at first insemination

	Model					
Predictor	1	2	3	4	5	6
Mid-infrared spectra					✓	✓
Cow information						
DIM, d	✓	✓		✓		✓
Parity, n	✓	✓		✓		✓
Calving month	✓	✓		✓		✓
Interval from calving to first insemination, d	✓	✓		✓		✓
Traditional milk information						
MY, kg/d		✓	✓	✓		✓
Fat, %		✓	✓	✓		
Protein, %		✓	✓	✓		
Lactose, %		✓	✓	✓		
SCC, 1,000/mL		✓	✓	✓		✓
Solids, %		✓	✓	✓		
Additional milk information						
BHB in blood, mmol/L			✓	✓		
MUN, mg/dL			✓	✓		
Milk freezing point depression, m°C			✓	✓		
MUFA, %			✓	✓		
PUFA, %			✓	✓		
SFA, %			✓	✓		
UFA, %			✓	✓		
C16:0, %			✓	✓		
C18:0, %			✓	✓		
C18:1n-6, %			✓	✓		

Table 2. Least squares means (± SE) for different variables in groups of cows with poor, average, and good likelihood of conception at first insemination[1]

Variable	Good (n = 1,745)	Average (n = 2,100)	Poor (n = 1,021)
DIM,[2] d	39.47 ± 0.66[a]	37.16 ± 0.65[b]	41.56 ± 0.66[a]
Interval from calving to first insemination,[2] d	69.07 ± 0.80[a]	69.61 ± 0.78[a]	79.24 ± 0.80[b]
Milk yield, kg/d	37.59 ± 0.30[a]	38.17 ± 0.31[ab]	38.95 ± 0.31[b]
Fat, %	4.32 ± 0.03[a]	4.23 ± 0.03[ab]	4.21 ± 0.03[b]
Protein, %	3.19 ± 0.01[a]	3.14 ± 0.01[b]	3.14 ± 0.01[b]
Lactose, %	5.19 ± 0.01[a]	5.16 ± 0.01[b]	5.15 ± 0.01[b]
SCC,[3] 1,000/mL	156.70 ± 12.84[a]	180.92 ± 12.57[a]	232.23 ± 12.91[b]
Solids, %	13.93 ± 1.6[a]	13.68 ± 0.05[b]	13.65 ± 0.05[b]
BHB in blood, mmol/L	1.65 ± 0.02[a]	1.64 ± 0.02[a]	1.58 ± 0.02[b]
MUN, mg/dL	13.77 ± 0.12[a]	13.39 ± 0.12[ab]	13.04 ± 0.12[b]
Freezing point depression, m°C	493.25 ± 3.77[a]	483.01 ± 3.69[a]	443.42 ± 3.57[b]
MUFA, %	0.99 ± 0.01[a]	1.03 ± 0.01[a]	1.03 ± 0.02[a]
PUFA, %	0.09 ± 0.002[a]	0.10 ± 0.002[a]	0.11 ± 0.002[b]
SFA, %	2.80 ± 0.03[a]	2.78 ± 0.03[a]	2.87 ± 0.03[a]
UFA, %	1.11 ± 0.02[a]	1.14 ± 0.02[a]	1.16 ± 0.02[a]
C16:0, %	1.40 ± 0.02[a]	1.38 ± 0.02[a]	1.43 ± 0.02[a]
C18:0, %	0.50 ± 0.01[a]	0.51 ± 0.01[a]	0.51 ± 0.01[a]
C18:1n-6, %	0.78 ± 0.01[a]	0.80 ± 0.01[a]	0.80 ± 0.01[a]

[a,b]For LC groups with the same letter, the difference between means was not significant ($P \geq 0.05$).

[1]Good = cows that conceived at first insemination; average = cows that conceived following 2 or more inseminations; poor = cows without conception event.

[2]Significance test performed on log-transformed values.

[3]Significance test performed on SCS.

reproduction data were merged with test-day and spectral data and edited. The initial CR ranged from 0.38 to 0.61 with an average of 0.47 and a SD of 0.06 across farms. The coefficient of variation for CR in the reproduction dataset was 13%. After data merging, quality control, and removal of spectral outliers, the average and good LC groups in our dataset represented around 79% of the records, while the average value (±SD) of CR decreased to 0.36 ± 0.08 and CR ranged from 0.21 to 0.52 in the remaining 3,451 cows. Significant differences in DIM, ICF, MY, and composition were observed between groups of cows with good, average, and poor LC (Table 2). Cows in the poor group had significantly higher ICF, MY, and SCC compared with cows in the good group, but lower concentrations of fat, protein, lactose, BHB, MUN,
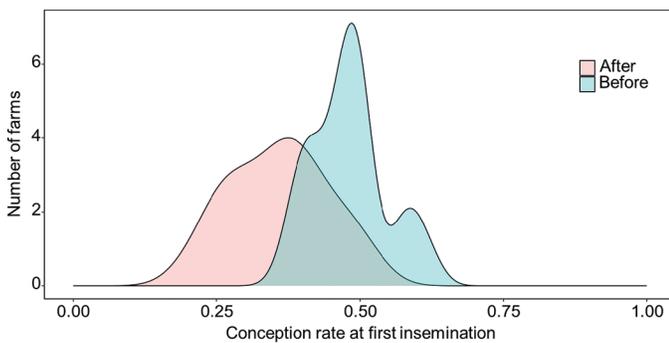

Figure 1. Distribution of conception rate at first insemination of Holstein cows across 18 farms before and after merging the data with available MIR spectra records and quality control.
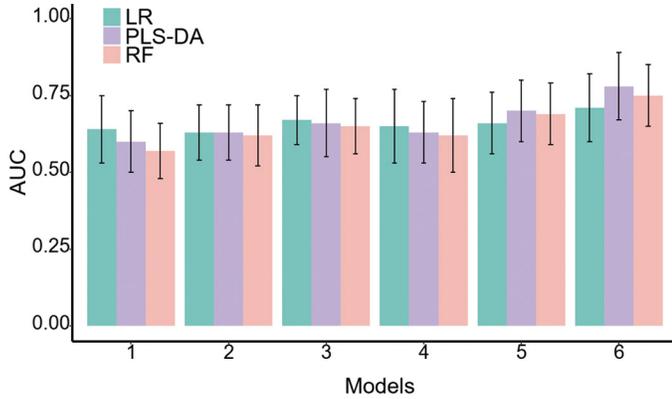
solids, and FPD. Fatty acid composition was invariant across groups, except for PUFA, which had a higher concentration in milk from the poor LC cows compared with average and good LC cows.

Comparison of Machine Learning Methods and Models

Table 3 shows the accuracies and AUC of the 6 models (models 1–6) when PLS-DA was used as the classification method. The prediction accuracy in random cross-validation was consistently slightly higher than that obtained in the herd-by-herd validation. Model 1, which used only cow information as predictors of LC, exhibited the lowest performance in the herd-by-herd validation (accuracy = 61.14% and AUC = 0.60). The average accuracy in the herd-by-herd validation improved by 2.35% by incorporating traditional milk information (model 2) and by 3.27% by using all milk information (model 3). Adding cow information to model 3 (model 4) increased the average accuracy to 67.44% in herd-by-herd validation. Using only MIR data (model 5) improved the model accuracy (+0.33% in random cross-validation and +0.12% in herd-by-herd validation) compared with model 4. Adding cow information, MY, and SCC to the MIR spectra (model 6) further increased the average accuracy by 8.79% in herd-by-herd validation. Particularly, model 6 consistently exhibited the highest sensitivity (73.14 ± 9.54%), and specificity (84.96 ± 11.10%) in herd-by-herd validation. Additionally, about 40% of the

Figure 2. Area under the curve (AUC) in the prediction of the likelihood of conception in herd-by-herd validation, using logistic regression (LR), partial least square discriminant analysis (PLS-DA), and random forest (RF) across 6 different models. The error bars indicate SD.

cows in the average LC group were incorrectly classified as poor, while the other 60% were classified as good by PLS-DA in model 6 in herd-by-herd validation.

As shown in Figure 2, the accuracy of LR and RF methods for the 6 models was similar to that of PLS-DA, with model 6 exhibiting the best results, followed by models 5 and 4. Logistic regression had better predictive performance in models with fewer predictors (models 1–4) compared with PLS-DA and RF. The accuracy of LR ranged from 63.21% to 69.59%, with AUC values ranging from 0.63 to 0.71 in herd-by-herd validation (Supplemental Table S1, see Notes). For models 5 and 6, the average prediction accuracies of RF in herd-by-herd validation were higher (67.93% and 75.30%, respectively) than those of LR (65.98% and 69.59%, respectively), but slightly lower than those obtained by PLS-DA (67.56% and 76.35%, respectively). The overall accuracy of RF ranged from 57.97% to 75.30%, with AUC ranging from 0.57 to 0.75 in the herd-by-herd validation (Supplemental Table S2, see Notes).

Supplemental Table S3 (see Notes) shows the negative control results obtained by LR, PLS-DA, and RF models, when only the water-related regions in MIR spectra were used as predictors. The accuracy of these models ranged from 47.97% to 52.59%, with AUC values ranging from 0.48 to 0.53 in herd-by-herd validation, respectively. The result of chi-squared test shows that their accuracies were not significantly different from 0.5 ($P \geq 0.05$).

Identification of the Most Influential Predictors for LC Classification

The VIP of MIR spectra wavenumbers and other predictors (DIM, parity, ICF, MY, and SCC) obtained for model 6 by PLS-DA in random cross-validation are shown in Figure 3. The predictors with the highest VIP

Table 3. Mean (± SD) accuracy and area under the curve (AUC) of the partial least square discriminant analysis models in classifying cows with poor or good likelihood of conception at first insemination[1]

Model	LV[2]	Random cross-validation				Herd-by-herd validation			
		Accuracy,[3] %	AUC	Specificity, %	Sensitivity, %	AUC	Accuracy, %	Specificity, %	Sensitivity, %
1	4	62.04 ± 5.27	0.62 ± 0.03	64.76 ± 5.76	59.51 ± 4.37	0.60 ± 0.10	61.14 ± 9.32	63.38 ± 10.61	57.40 ± 11.45
2	10	64.18 ± 4.98	0.64 ± 0.02	66.67 ± 3.86	62.71 ± 5.33	0.63 ± 0.09	63.49 ± 10.17	64.96 ± 10.66	60.11 ± 10.24
3	16	66.71 ± 3.78	0.66 ± 0.02	67.56 ± 3.91	64.67 ± 4.67	0.65 ± 0.10	64.41 ± 11.09	66.31 ± 10.84	63.28 ± 10.89
4	20	68.84 ± 6.12	0.69 ± 0.05	71.19 ± 5.78	67.86 ± 6.12	0.68 ± 0.11	67.44 ± 8.63	70.83 ± 10.63	66.13 ± 10.01
5	14	69.17 ± 5.78	0.68 ± 0.04	71.95 ± 5.88	68.26 ± 6.96	0.67 ± 0.10	67.56 ± 12.67	71.34 ± 10.73	66.87 ± 9.97
6	20	79.87 ± 6.49	0.80 ± 0.06	85.88 ± 6.99	74.43 ± 6.78	0.77 ± 0.11	76.35 ± 10.60	84.96 ± 11.10	73.14 ± 9.54

[1]Good = cows that conceived at first insemination; poor = cows without conception event.

[2]LV = number of latent variables (principal components).

[3]Accuracy = (TP + TN)/(TP + TN + FP + FN); sensitivity = TP/(TP + FN); specificity = TN/(TN + FP). TP, TN, FP, and FN represent the total numbers of true positives, true negatives, false positives, and false negatives as predicted by the model, respectively.
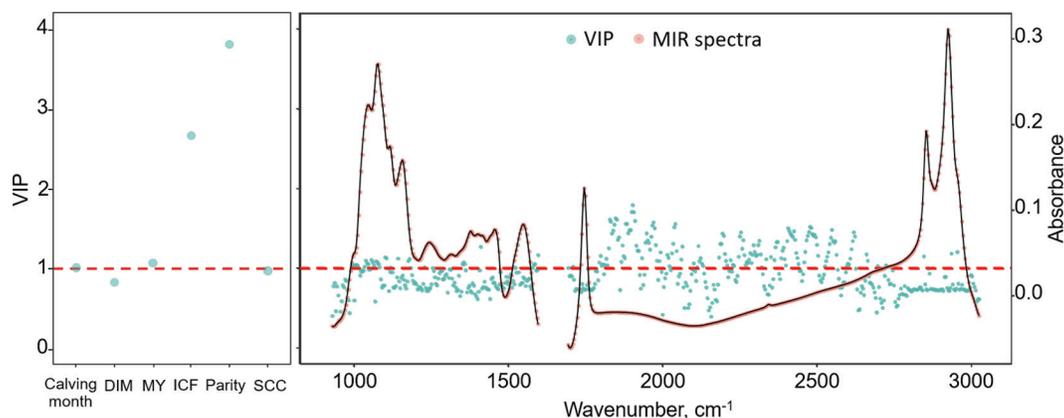
Figure 3. Variable importance in projection (VIP) obtained by partial least square discriminant analysis and random cross-validation for the predictors included in model 6. MY = milk yield; ICF = interval from calving to first insemination. The red dashed line indicates the threshold beyond which variables are considered to substantially contribute to the prediction (VIP > 1), and the blue dots indicate the VIP for the predictors.

were parity (VIP = 3.80) and ICF (VIP = 2.66), whereas SCC, MY, and DIM had VIP values around 1. The wavenumber with VIP ≥1.00 were located within 3 regions, namely from 1,003 to 1,189, 1,794 to 2,260, and 2,300 to 2,660 cm^{-1}, whereas the highest VIP values were found mainly from 1,827 to 1,970 cm^{-1}.

DISCUSSION

Data Editing and Selection

This research explores the potential of applying MIR spectra to predict LC in highly productive TMR systems. In comparison to prior research by Ho et al. (2019) and Ho and Pryce (2020), our study specifically used data from Holstein cows in China, whereas Holstein-Friesian, purebred Jersey, and crossbred cows in Australia constituted their sample. Up to 83% of Australian farms employ the grass-based or moderate-high bale system (Dairy Australia, 2020), whereas Chinese farms predominantly use the TMR system. In addition, there are notable difference in farm sizes between the 2 studies. Furthermore, variations in production and reproduction performances exist among the cows in the respective studies (i.e., our study incorporates an average annual MY per cow ranging from 12 to 13 t and a conception rate of 0.38 to 0.61). Ho et al. (2019) report an average annual MY of ~7 t/cow, with a conception rate ranging from 0.22 to 0.54.

Data on cows from different farms were used to ensure sufficient representativeness and maximize the variability in CR (Figure 1), as data diversity determines the robustness of prediction models. The 13% coefficient of variation for CR can be considered satisfactory as variation in farm-level fertility may result from differences in management, farm size, and genetics (Chebel et al., 2004; Grimard et al., 2006), whereas most of the farms

included in the study had similar reproduction management. The mean CR (0.47 ± 0.06) was similar to that observed in a previous study on Chinese Holstein cows (around 55%; Guo et al., 2014). However, CR used in prediction after data editing was on average 0.36 (±0.08; Figure 1). The difference between the average CR before and after data editing indicates that in particular the cows with good LC (and thus high CR) were excluded from the study because of missing MIR spectra. This is because cows with good LC have shorter ICF and might have less opportunity to obtain MIR spectra before the first insemination than cows with average and poor LC (i.e., within the first 2 mo). In particular, the comparison between the ICF of the good and poor LC group (Table 2) indicates that the estrus in the good LC group resumed earlier after calving compared with the poor LC group. This result can be explained by the management practices in Chinese farms. Cows with a low BCS or a disease, that are thought to have a poor LC, will be synchronized and inseminated at a later date after calving.

The average and good LC groups in our dataset represented around 79% of the records. This was expected, as field datasets are often predominantly composed of "normal" observations with only a small percentage of "abnormal" or "attractive" records (i.e., cows with poor LC). In this case, traditional machine learning algorithms may tend to favor the dominant category, resulting in poor model performance when the aim is to predict the under-represented category. Synthetic minority over-sampling techniques have been proposed to address the issue before modeling by over-sampling the minority (abnormal) category or under-sampling the majority (normal) category to balance the dataset (Chawla et al., 2002). The elimination of the average group in the training process was inspired by SMOTE and used in this study to reduce the tendency of models to predict the

most frequent category and increase the discrimination of extreme groups' features. In a preliminary analysis, we attempted to train 3-class classification models using the good, average, and poor groups (Supplemental Table S4, see Notes). Random forest showed the best performance across methods in the herd-by-herd validation, with an accuracy (±SD) equal to 49.21% ± 8.41%, a sensitivity of 51.92% ± 10.91%, and a specificity of 61.67% ± 11.42%, but the performance was worse than the one of the corresponding 2-class classification models (accuracy = 0.75, Supplemental Table S2). Hence, our results confirm the benefits of SMOTE. Likewise, studies in humans indicated that the SMOTE method effectively improved risk gene discovery, detection of fatal congenital heart disease and of spinal disease (Fernández et al., 2018; Sowjanya and Mrudula, 2023; Zhang et al., 2023). The method has previously been used in cows in the detection of insemination problems (Grzesiak et al., 2010) and LC prediction (Ho et al., 2019).

Comparison of Machine Learning Methods and Models

The 3 machine learning methods had similar performance in predicting LC in dairy cows (Figure 2). However, PLS-DA facilitates the identification of system information and noise, particularly for explanatory predictors with large multicollinearity (e.g., milk MIR-spectra). Logistic regression, in contrast, exhibited poor performance when the MIR wavenumbers were used as predictors (Figure 2). This outcome was expected, as LR struggles with handling a high number of features or variables, and it is susceptible to overfitting (McCullagh, 1984). A similar finding was also reported in the prediction of conception success to a given insemination in lactating dairy cows, where the average AUC using only MIR spectra was 0.57, which is lower than 0.66 obtained when the MIR spectra were not included in the models (Hempstalk et al., 2015). Hence, PLS-DA and RF have the greatest potential to become the methods of choice for the prediction of LC.

The values of AUC in herd-by-herd validation obtained in this study, ranging from 0.60 to 0.77, slightly surpassed those (0.48–0.75) reported by Hempstalk et al. (2015), Ho et al. (2019), and Ho and Pryce (2020), but were lower than the AUC (0.91) reported by Grzesiak et al. (2010). Compared with the AUC obtained in models from 1 to 4, the AUC from model 6 (0.77 in the herd-by-herd validation) is to be considered good (Šimundić, 2009), indicating that milk MIR spectra can capture differences in metabolic or reproductive status in cows, as previously suggested (Grelet et al., 2019; Luke et al., 2019). Notably, the specificity (ranging from 74.4%–84.9%) was higher than the sensitivity (from

67.8%–73.1%), which implies that the model has a stronger ability to identify a cow with poor LC than a cow with good LC. This is an advantage, as it allows us to identify the cows that might need treatments to improve the CR. In practice, the dairy population normally includes good, average, or poor LC cows on farms. Based on the 2-classification model in this study, the predictive ability for average group was lost, as they were not used in model training, and approximately 40% of the cows with an average LC were classified as having a poor LC. These additional individuals predicted as poor LC cows might also deserve farmers' attention, because they might be in a similar metabolic status to that of the true poor LC cows and may require more inseminations per pregnancy, with a potential negative effect on the farm.

As expected, in all models the average prediction accuracy in random cross-validation was consistently higher than that of the herd-by-herd validation (Table 3), in line with previous reports (Ho et al., 2019; Luke et al., 2019). Despite the encouraging results, the higher SD of prediction accuracy obtained in the herd-by-herd validation indicated that performances were variable across farms, and thus, a more extensive dataset is needed to improve the robustness of the model. The accuracy obtained for the models tested as negative control (using only the water-related regions of the MIR spectra as predictors of LC) was not different from 0.5 ($P \geq 0.05$), which corresponds to a random classification of the cows across LC groups. All data were collected from instruments of the same brand, calibrated monthly in the same laboratory, and this result confirms that factors such as instrument, batch, and time did not affect LC predictions, thereby establishing the credibility of our estimates.

Furthermore, in a preliminary analysis we also attempted to identify the milk testing period between calving and first insemination that maximized the prediction accuracy of LC. In the herd-by-herd validation, spectra collected in proximity to the first insemination exhibited slightly superior prediction accuracies compared with those collected after calving (results not reported in tables). However, variations in accuracy and specificity across time periods were minimal. A future comparison of the prediction accuracy of models based on daily MIR spectra collected before the first insemination might enable the determination of the optimal time of milk testing, provided sufficient spectra are available. This might facilitate the future application of the model and reduce sampling costs.

Best Predictors of the LC

Parity (VIP = 3.80) and ICF (VIP = 2.66) were the variables that mostly influenced the prediction of LC. Both of them have been previously associated with CR in

cows (Buckley et al., 2003; Berry et al., 2011). Published results indicated that a low and abnormal progesterone concentration after first insemination exists in the second and third lactations (Hommeida et al., 2004). Ovarian recrudescence with regular re-occurring estrous cycles and fertility restoration to first service is associated with a reduced occurrence of periparturient metabolic and reproductive disorders (Thatcher et al., 2006), which are more likely to manifest as the number of parities increases. This means that cows with a higher parity have a longer ICF, as evidenced by the cows in the good group having shorter ICF than the poor group (Table 2).

Three spectra regions (in the ranges 1,003–1,189, 1,794–2,260, and 2,300–2,660 cm^{-1}) were identified in our study as associated with LC prediction (Figure 3). The same regions were found to be associated with LC by Ho and Pryce (2020), whereas Toledo-Alvarado et al. (2018a) found associations with pregnancy status at around 1,550 cm^{-1}. From a chemical point of view, these regions are mainly related to complex chemical bonds including those of alkenes (–CH2, –CH3, and C=C), aromatic or aliphatic compounds (C–H, C=C, and C–C), alcohols (C–O and –OH), ethers (C–O), amines and amides (C–N, N–N, and N–H), aldehydes and ketones (C=O), carboxyl groups (–OH), nitrates, and urea (Duffy, 1972). These chemical bonds are probably related to fat, protein, ketone bodies, milk components, sex hormones, and other substances, as indicated by the significant differences in fat, protein, lactose, SCC, BHB, and MUN between the good and poor LC group (Table 2).

A relationship between fertility and milk composition has been reported by several studies. Morton et al. (2017) found that cows with a higher milk protein content in early lactation had a substantially better reproductive performance, whereas Toledo-Alvarado et al. (2018b) observed differences in milk traits (particularly fat, protein, and lactose) during the 10 d before insemination between cows that conceived or did not conceive. Other studies have demonstrated that cows that develop clinical mastitis in early lactation are less likely to conceive, especially around insemination (Santos et al., 2004; Lavon et al., 2011), due to the release of inflammatory mediators such as $PGF_{2\alpha}$, which can alter the interestrus interval by causing premature luteolysis (Hertl et al., 2010). The magnitude or rate of mobilization of body fat and protein usually leads to a higher concentration of blood BHB and MUN within a normal range. The slightly higher BHB and MUN concentration observed in the good LC group indicates a higher energy expenditure from the good LC group to maintain good performance such as reproduction compared with the poor LC group, despite the average BHB of cows >1.2 mmol/L suggesting that they may experience subclinical ketosis (Butler, 2000; Lou et al., 2022).

There seems to be a complex relationship among reproduction, milk performance, and MIR spectra during early lactation. The results of this study are insufficient to fully explain these intricacies, and only confirm the association between successful and unsuccessful reproduction (i.e., low CR at first insemination) and metabolic regulation (i.e., milk components) manifested in the phenotype (i.e., milk MIR spectra). Based on easily collected MIR spectra, the prediction model developed in this study could be included in decision support tools to help farmers optimize their mating decisions in advance, saving time as well as additional costs of semen and wages of veterinarians. Furthermore, it could be also a potential indicator trait in dairy cattle breeding. The large-scale availability of MIR spectra and the availability of LC prediction models as developed in this study could address the challenge posed by the limited accessibility of reproduction phenotypes due to their difficult collection, and enable genetic and genomic analyses for reproduction traits when phenotype collection is not yet routinely established.

CONCLUSIONS

Milk MIR spectra collected in early lactation and in combination with basic on-farm data could be used as input in a model with PLS-DA to identify cows with good and poor LC with 76% accuracy. The model had a superior ability to identify cows with poor LC compared with cows with good LC. Application of such as model could help farmers to make decisions before insemination and effectively improve the performance of cows with poor LC, which could save costs. In addition, parity, ICF, and 3 MIR regions (1,003–1,189, 1,794–2,260, and 2,300–2,660 cm^{-1}) were the predictors mostly associated with LC at first insemination. This study further confirms the potential of applying MIR spectra in reproduction management in highly productive TMR systems. More studies should explore the relationship between reproduction and MIR spectra as more data accumulate.

NOTES

.26708542.v1 (Supplemental Table S2), https://doi.org/10.6084/m9.figshare.26708614.v1 (Supplemental Table S3), and https://doi.org/10.6084/m9.figshare.26708668.v1 (Supplemental Table S4). Milk sample collection at the farms was conducted under the Dairy Herd Improvement program; therefore, no ethical approval (Institutional Animal Care and Use Committee) was required for this study. The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: AUC = area under the curve; CR = calving rate; FP = fat percentage; FPD = freezing point depression; GH = global H distance; ICF = interval from calving to first; LC = likelihood of conception; LP = lactose percentage; LR = logistic regression; MIR = mid-infrared; MY = milk yield; PC = principal component; PLS-DA = partial least square discriminant analysis; PP = protein percentage; RF = random forest; SMOTE = synthetic minority over-sampling technique; VIP = variable importance in projection.

REFERENCES

Benedet, A., M. Franzoi, M. Penasa, E. Pellattiero, and M. De Marchi. 2019. Prediction of blood metabolites from milk mid-infrared spectra in early-lactation cows. J. Dairy Sci. 102:11298–11307. https://doi.org/10.3168/jds.2019-16937.

Berry, D. P., R. D. Evans, and S. McParland. 2011. Evaluation of bull fertility in dairy and beef cattle using cow field data. Theriogenology 75:172–181. https://doi.org/10.1016/j.theriogenology.2010.08.002.

Blavy, P., N. C. Friggens, K. R. Nielsen, J. M. Christensen, and M. Derks. 2018. Estimating probability of insemination success using milk progesterone measurements. J. Dairy Sci. 101:1648–1660. https://doi.org/10.3168/jds.2016-12453.

Bonfatti, V., P. N. Ho, and J. E. Pryce. 2020. Usefulness of milk mid-infrared spectroscopy for predicting lameness score in dairy cows. J. Dairy Sci. 103:2534–2544. https://doi.org/10.3168/jds.2019-17551.

Brand, W., A. T. Wells, S. L. Smith, S. J. Denholm, E. Wall, and M. P. Coffey. 2021. Predicting pregnancy status from mid-infrared spectroscopy in dairy cow milk using deep learning. J. Dairy Sci. 104:4980–4990. https://doi.org/10.3168/jds.2020-18367.

Breiman, L. 2001. Random forests. Mach. Learn. 45:5–32. https://doi.org/10.1023/A:1010933404324.

Buckley, F., J. Mee, K. O'Sullivan, R. Evans, D. Berry, and P. Dillon. 2003. Insemination factors affecting the conception rate in seasonal calving Holstein-Friesian cows. Reprod. Nutr. Dev. 43:543–555. https://doi.org/10.1051/rnd:2004002.

Butler, W. R. 2000. Nutritional interactions with reproductive performance in dairy cattle. Anim. Reprod. Sci. 60–61:449–457. https://doi.org/10.1016/S0378-4320(00)00076-2.

Cabrera, V. E. 2014. Economics of fertility in high-yielding dairy cows on confined TMR systems. Animal 8:211–221. https://doi.org/10.1017/S1751731114000512.

Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16:321–357. https://doi.org/10.1613/jair.953.

Chebel, R. C., J. E. P. Santos, J. P. Reynolds, R. L. A. Cerri, S. O. Juchem, and M. Overton. 2004. Factors affecting conception rate after artificial insemination and pregnancy loss in lactating dairy cows. Anim. Reprod. Sci. 84:239–255. https://doi.org/10.1016/j.anireprosci.2003.12.012.

Dairy Australia. 2020. A guide for investment and the Australian dairy industry. Accessed Mar. 25, 2024. https://cdn-prod.dairyaustralia.com.au/-/media/project/dairy-australia-sites/national-home/resources/2020/07/09/guide-to-investment-and-australian-dairy-industry/guide-to-investment-and-australian-dairy-industry.pdf?rev=c0b0b8c2ba0448c7a4b4df7253949850.

De Marchi, M., V. Toffanin, M. Cassandro, and M. Penasa. 2014. Invited review: Mid-infrared spectroscopy as phenotyping tool for milk traits. J. Dairy Sci. 97:1171–1186. https://doi.org/10.3168/jds.2013-6799.

DeJarnette, J. M., C. E. Marshall, R. W. Lenz, D. R. Monke, W. H. Ayars, and C. G. Sattler. 2004. Sustaining the fertility of artificially inseminated dairy cattle: The role of the artificial insemination industry. J. Dairy Sci. 87:E93–E104. https://doi.org/10.3168/jds.S0022-0302(04)70065-X.

Duffy, N. V. 1972. Interpretation of infrared spectra. J. Chem. Educ. 49:652. https://doi.org/10.1021/ed049p652.1.

Fernández, A., S. García, F. Herrera, and N. V. Chawla. 2018. SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. J. Artif. Intell. Res. 61:863–905. https://doi.org/10.1613/jair.1.11192.

Franzoi, M., G. Niero, G. Meoni, L. Tenori, C. Luchinat, M. Penasa, M. Cassandro, and M. De Marchi. 2023. Effectiveness of mid-infrared spectroscopy for the prediction of cow milk metabolites. J. Dairy Sci. 106:5288–5297. https://doi.org/10.3168/jds.2023-23226.

Grelet, C., C. Bastin, M. Gelé, J. B. Davière, M. Johan, A. Werner, R. Reding, J. A. F. Fernandez Pierna, F. G. Colinet, P. Dardenne, N. Gengler, H. Soyeurt, and F. Dehareng. 2016. Development of Fourier transform mid-infrared calibrations to predict acetone, β-hydroxybutyrate, and citrate contents in bovine milk through a European dairy network. J. Dairy Sci. 99:4816–4825. https://doi.org/10.3168/jds.2015-10477.

Grelet, C., P. Dardenne, H. Soyeurt, J. A. Fernandez, A. Vanlierde, F. Stevens, N. Gengler, and F. Dehareng. 2021. Large-scale phenotyping in dairy sector using milk MIR spectra: Key factors affecting the quality of predictions. Methods 186:97–111. https://doi.org/10.1016/j.ymeth.2020.07.012.

Grelet, C., J. A. Fernández Pierna, P. Dardenne, V. Baeten, and F. Dehareng. 2015. Standardization of milk mid-infrared spectra from a European dairy network. J. Dairy Sci. 98:2150–2160. https://doi.org/10.3168/jds.2014-8764.

Grelet, C., A. Vanlierde, M. Hostens, L. Foldager, M. Salavati, M. Ingvartsen, M. Crowe, M. T. Sorensen, E. Froidmont, C. P. Ferris, C. Marchitelli, F. Becker, T. Larsen, F. Carter, and F. Dehareng. 2019. Potential of milk mid-IR spectra to predict metabolic status of cows through blood components and an innovative clustering approach. Animal 13:649–658. https://doi.org/10.1017/S1751731118001751.

Grimard, B., S. Freret, A. Chevallier, A. Pinto, C. Ponsart, and P. Humblot. 2006. Genetic and environmental factors influencing first service conception rate and late embryonic/foetal mortality in low fertility dairy herds. Anim. Reprod. Sci. 91:31–44. https://doi.org/10.1016/j.anireprosci.2005.03.003.

Gröhn, Y. T., and P. J. Rajala-Schultz. 2000. Epidemiology of reproductive performance in dairy cows. Anim. Reprod. Sci. 60–61:605–614. https://doi.org/10.1016/S0378-4320(00)00085-3.

Grzesiak, W., D. Zaborski, P. Sablik, A. Zukiewicz, A. Dybus, and I. Szatkowska. 2010. Detection of cows with insemination problems using selected classification models. Comput. Electron. Agric. 74:265–273. https://doi.org/10.1016/j.compag.2010.09.001.

Guo, G., X. Guo, Y. Wang, X. Zhang, S. Zhang, X. Li, L. Liu, W. Shi, T. Usman, X. Wang, L. Du, and Q. Zhang. 2014. Estimation of genetic parameters of fertility traits in Chinese Holstein cattle. Can. J. Anim. Sci. 94:281–285. https://doi.org/10.4141/cjas2013-113.

Hempstalk, K., S. Mcparland, and D. P. Berry. 2015. Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows. J. Dairy Sci. 98:5262–5273. https://doi.org/10.3168/jds.2014-8984.

Hertl, J. A., Y. T. Gröhn, J. D. G. Leach, D. Bar, G. J. Bennett, R. N. González, B. J. Rauch, F. L. Welcome, L. W. Tauer, and Y. H. Schukken. 2010. Effects of clinical mastitis caused by gram-positive and gram-negative bacteria and other organisms on the probability of conception in New York State Holstein dairy cows. J. Dairy Sci. 93:1551–1560. https://doi.org/10.3168/jds.2009-2599.

Ho, P. N., T. D. W. Luke, and J. E. Pryce. 2021. Validation of milk mid-infrared spectroscopy for predicting the metabolic status of lactating dairy cows in Australia. J. Dairy Sci. 104:4467–4477. https://doi.org/10.3168/jds.2020-19603.

Ho, P. N., V. Bonfatti, T. D. W. Luke, and J. E. Pryce. 2019. Classifying the fertility of dairy cows using milk mid-infrared spectroscopy. J. Dairy Sci. 102:10460–10470. https://doi.org/10.3168/jds.2019-16412.

Ho, P. N., and J. E. Pryce. 2020. Predicting the likelihood of conception to first insemination of dairy cows using milk mid-infrared spectroscopy. J. Dairy Sci. 103:11535–11544. https://doi.org/10.3168/jds.2020-18589.

Hommeida, A., T. Nakao, and H. Kubota. 2004. Luteal function and conception in lactating cows and some factors influencing luteal function after first insemination. Theriogenology 62:217–225. https://doi.org/10.1016/j.theriogenology.2003.09.018.

Kaniyamattam, K., M. A. Elzo, J. B. Cole, and A. De Vries. 2016. Stochastic dynamic simulation modeling including multitrait genetics to estimate genetic, technical, and financial consequences of dairy farm reproduction and selection strategies. J. Dairy Sci. 99:8187–8202. https://doi.org/10.3168/jds.2016-11136.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28. https://doi.org/10.18637/jss.v028.i05.

Lavon, Y., E. Ezra, G. Leitner, and D. Wolfenson. 2011. Association of conception rate with pattern and level of somatic cell count elevation relative to time of insemination in dairy cows. J. Dairy Sci. 94:4538–4545. https://doi.org/10.3168/jds.2011-4293.

Lou, W., H. Zhang, H. Luo, Z. Chen, R. Shi, X. Guo, Y. Zou, L. Liu, L. F. Brito, G. Guo, and Y. Wang. 2022. Genetic analyses of blood β-hydroxybutyrate predicted from milk infrared spectra and its association with longevity and female reproductive traits in Holstein cattle. J. Dairy Sci. 105:3269–3281. https://doi.org/10.3168/jds.2021-20389.

Luke, T. D. W., S. Rochfort, W. J. Wales, V. Bonfatti, L. Marett, and J. E. Pryce. 2019. Metabolic profiling of early-lactation dairy cows using milk mid-infrared spectra. J. Dairy Sci. 102:1747–1760. https://doi.org/10.3168/jds.2018-15103.

McCullagh, P. 1984. Generalized linear models. Eur. J. Oper. Res. 16:285–292. https://doi.org/10.1016/0377-2217(84)90282-0.

Mehmood, T., K. H. Liland, L. Snipen, and S. Sæbø. 2012. A review of variable selection methods in Partial Least Squares Regression. Chemom. Intell. Lab. Syst. 118:62–69. https://doi.org/10.1016/j.chemolab.2012.07.010.

Morton, J. M., M. J. Auldist, M. L. Douglas, and K. L. Macmillan. 2017. Milk protein concentration, estimated breeding value for fertility, and reproductive performance in lactating dairy cows. J. Dairy Sci. 100:5850–5762. https://doi.org/10.3168/jds.2016-11273.

Mullen, K. M., and I. H. M. Stokkum. 2007. An introduction to the "Special Volume Spectroscopy and Chemometrics in R". J. Stat. Softw. 18. https://doi.org/10.18637/jss.v018.i01.

Rearte, R., S. J. Leblanc, S. G. Corva, R. L. de la Sota, I. M. Lacau-Mengido, and M. J. Giuliodori. 2018. Effect of milk production on reproductive performance in dairy herds. J. Dairy Sci. 101:7575–7584. https://doi.org/10.3168/jds.2017-13796.

Santos, J. E. P., R. L. A. Cerri, M. A. Ballou, G. E. Higginbotham, and J. H. Kirk. 2004. Effect of timing of first clinical mastitis occurrence on lactational and reproductive performance of Holstein dairy cows. Anim. Reprod. Sci. 80:31–45. https://doi.org/10.1016/S0378-4320(03)00133-7.

Savitzky, A., and M. J. E. Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36:1627–1639. https://doi.org/10.1021/ac60214a047.

Shahinfar, S., D. Page, J. Guenther, V. Cabrera, P. Fricke, and K. Weigel. 2014. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. J. Dairy Sci. 97:731–742. https://doi.org/10.3168/jds.2013-6693.

Šimundić, A.-M. 2009. Measures of diagnostic accuracy: Basic definitions. EJIFCC 19:203–211.

Sowjanya, A. M., and O. Mrudula. 2023. Effective treatment of imbalanced datasets in health care using modified SMOTE coupled with stacked deep learning algorithms. Appl. Nanosci. 13:1829–1840. https://doi.org/10.1007/s13204-021-02063-4.

Thatcher, W. W., T. R. Bilby, J. A. Bartolome, F. Silvestre, C. R. Staples, and J. E. P. Santos. 2006. Strategies for improving fertility in the modern dairy cow. Theriogenology 65:30–44. https://doi.org/10.1016/j.theriogenology.2005.10.004.

Toledo-Alvarado, H., A. I. Vazquez, G. de los Campos, R. J. Tempelman, G. Bittante, and A. Cecchinato. 2018a. Diagnosing pregnancy status using infrared spectra and milk composition in dairy cows. J. Dairy Sci. 101:2496–2505. https://doi.org/10.3168/jds.2017-13647.

Toledo-Alvarado, H., A. I. Vazquez, G. de los Campos, R. J. Tempelman, G. Gabai, A. Cecchinato, and G. Bittante. 2018b. Changes in milk characteristics and fatty acid profile during the estrous cycle in dairy cows. J. Dairy Sci. 101:9135–9153. https://doi.org/10.3168/jds.2018-14480.

van den Berg, I., P. N. Ho, M. Haile-Mariam, and J. E. Pryce. 2021. Genetic parameters for mid-infrared spectroscopy–predicted fertility. JDS Commun. 2:361–365. https://doi.org/10.3168/jdsc.2021-0141.

Vanrobays, M. L., C. Bastin, J. Vandenplas, H. Hammami, H. Soyeurt, A. Vanlierde, F. Dehareng, E. Froidmont, and N. Gengler. 2016. Changes throughout lactation in phenotypic and genetic correlations between methane emissions and milk fatty acid contents predicted from milk mid-infrared spectra. J. Dairy Sci. 99:7247–7260. https://doi.org/10.3168/jds.2015-10646.

Wang, Q., and H. Bovenhuis. 2019. Validation strategy can result in an overoptimistic view of the ability of milk infrared spectra to predict methane emission of dairy cattle. J. Dairy Sci. 102:6288–6295. https://doi.org/10.3168/jds.2018-15684.

Zhang, B., X. Dong, Y. Hu, X. Jiang, and G. Li. 2023. Classification and prediction of spinal disease based on the SMOTE-RFE-XGBoost model. PeerJ Comput. Sci. 9:e1280. https://doi.org/10.7717/peerj-cs.1280.

ORCIDS

W. Lou ⓘ https://orcid.org/0000-0002-0150-6068

V. Bonfatti ⓘ https://orcid.org/0000-0003-3970-5764

H. Bovenhuis ⓘ https://orcid.org/0000-0002-9074-5334

R. Shi ⓘ https://orcid.org/0000-0002-0047-1262

A. van der Linden ⓘ https://orcid.org/0000-0002-2363-356X

H. A. Mulder ⓘ https://orcid.org/0000-0003-2124-4787

Y. Wang ⓘ https://orcid.org/0000-0003-3629-2802

B. Ducro ⓘ https://orcid.org/0000-0002-0866-0199