Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods

M. Frizzarin,[1,2] **I. C. Gormley,**[1]* **D. P. Berry,**[2] **T. B. Murphy,**[1] **A. Casa,**[1] **A. Lynch,**[1]
and S. McParland[2]*†
[1]School of Mathematics and Statistics, University College Dublin, Belfield, Dublin 4, Ireland
[2]Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, P61 P302 Ireland

ABSTRACT

Numerous statistical machine learning methods suitable for application to highly correlated features, as those that exist for spectral data, could potentially improve prediction performance over the commonly used partial least squares approach. Milk samples from 622 individual cows with known detailed protein composition and technological trait data accompanied by mid-infrared spectra were available to assess the predictive ability of different regression and classification algorithms. The regression-based approaches were partial least squares regression (PLSR), ridge regression (RR), least absolute shrinkage and selection operator (LASSO), elastic net, principal component regression, projection pursuit regression, spike and slab regression, random forests, boosting decision trees, neural networks (NN), and a post-hoc approach of model averaging (MA). Several classification methods (i.e., partial least squares discriminant analysis (PLSDA), random forests, boosting decision trees, and support vector machines (SVM)) were also used after stratifying the traits of interest into categories. In the regression analyses, MA was the best prediction method for 6 of the 14 traits investigated [curd firmness at 60 min, α_{S1}-casein (CN), α_{S2}-CN, κ-CN, α-lactalbumin, and β-lactoglobulin B], whereas NN and RR were the best algorithms for 3 traits each (rennet coagulation time, curd-firming time, and heat stability, and curd firmness at 30 min, β-CN, and β-lactoglobulin A, respectively), PLSR was best for pH, and LASSO was best for CN micelle size. When traits were divided into 2 classes, SVM had the greatest accuracy for the majority of the traits investigated. Although the well-established PLSR-based method performed competitively, the application of statistical machine learning methods for regression analyses reduced the root mean square error compared with PLSR from between 0.18% (κ-CN) to 3.67% (heat stability). The use of modern statistical machine learning methods for trait prediction from mid-infrared spectroscopy may improve the prediction accuracy for some traits.

Key words: Fourier-transform mid-infrared spectroscopy, statistical machine learning, milk quality

INTRODUCTION

Fourier-transform mid-infrared spectroscopy (**MIRS**) is a methodology that exploits mid-infrared region light to indirectly predict the concentration of constituents in a sample. When a sample is analyzed using MIRS, light is passed through the sample at a sequence of wavelengths in the mid-infrared region (5,000 to 900 cm^{-1}), activating the chemical bonds of the sample matter with a consequential effect on the absorption of energy from the light (Skolik et al., 2018). The extent of the energy absorbed creates the spectrum for that sample, which should therefore be useful to predict the quantity of individual components within the sample. Infrared spectroscopy is used in different fields, from medicine (Petrich, 2001) to astronomy (Keller et al., 2006), as well as in animal science (De Marchi et al., 2014).

Mid-infrared spectroscopy is a low-cost, rapid and nondisruptive technique, routinely used in the analysis of cow milk samples for the determination of fat, protein, lactose, and CN concentrations in both bulk and individual animal samples (De Marchi et al., 2014). For this reason, MIRS is a potentially useful vehicle for collecting vast quantities of data at a population level. The literature documents the ability of MIRS to predict novel milk-related traits such as the coagulation properties of milk (Cecchinato et al., 2009; Visentin et al., 2016; El Jabri et al., 2019) and individual milk fatty acids (Soyeurt et al., 2006; Bonfatti et al., 2017), as well as animal-related traits, such as energy efficiency (McParland et al., 2014), energy intake (McParland and Berry, 2016), and methane emissions (Dehareng et al., 2012).

Partial least squares regression (**PLSR**) has traditionally been the method of choice in relating MIRS data of cow milk to novel milk and animal characteristics, owing to its capability to consider collinear, high-dimensional data sets. Nonetheless, the investigation of the application of other statistical machine learning (**ML**) methods in predicting an outcome variable has been demonstrated in animal science research. Both Li et al. (2018) and Xu et al. (2019) used novel ML methods to respectively predict phenotypic performance using SNP data (Li et al., 2018) or cow metabolic status from animal and herd-level features. Nevertheless, the application of statistical ML methods in MIRS analyses is still rare. The potential usefulness of statistical ML methods to predict milk traits from spectra is due to their ability to perform well in multidimensional correlated data but also importantly to identify nonlinear associations between the wavelengths and the observed value of the trait. Recently, the division of continuous traits into categories before MIRS prediction analyses has also been considered (Manuelian et al., 2017; Grelet et al., 2019; Duplessis et al., 2020).

The novelty of the present study is the evaluation of modern statistical ML methods in predicting a series of cow milk quality traits, including milk technological traits (i.e., rennet coagulation time, curd-firming time, curd firmness at 30 and 60 min, CN micelle size, pH, and heat stability) and individual milk proteins (i.e., α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA, β-LG A, and β-LG B) from milk MIRS. These outcome traits were also divided into categories and the performance of modern classification methods assessed with the purpose of determining which performs best. The use of modern statistical ML methods for trait prediction from MIRS may improve the prediction accuracy for some traits.

MATERIALS AND METHODS

Data

The data set used in the present study is described in detail by both Visentin et al. (2015) and McDermott et al. (2016). In brief, 730 milk samples from 622 cows were collected between August 2013 and August 2014 from 7 different Irish research herds. The samples originated from Holstein-Friesian, Jersey, and Norwegian Red cows, as well as their crosses; all cows were fed a predominantly grass-based diet with occasional concentrate and grass silage supplementation. The samples were collected during morning and evening milking and represented different stages of lactation and different parities. All samples were analyzed by the same MilkoScan FT6000 (Foss Electronic A/S), and

the resulting spectrum, comprising 1,060 transmittance data points in the mid-infrared light region, was stored. The traits investigated in the present study included the milk technological traits of rennet coagulation time (**RCT**), curd-firming time (**k20**), curd firmness at 30 and 60 min (**a30**, **a60**), casein micelle size (**CMS**), pH, and heat stability, as well as detailed milk protein traits, including α_{S1}-CN, α_{S2}-CN, β-CN, κ-CN, α-LA, β-LG A, and β-LG B.

The milk coagulation properties were quantified using a Formagraph (Foss Electronic A/S). Milk pH of all samples was assessed with a SevenCompact pH meter S220 (Mettler Toledo AG). The CN micelle hydrodynamic diameter was determined using a Zetasizer Nano system (Malvern Instruments Inc.). Heat stability was tested using the method outlined by Davies and White (1966). Milk proteins were determined using reverse-phase HPLC using an adaptation of the method of Visser et al. (1991) and are expressed as grams per liter of milk.

To satisfy the assumption that all observations used were independent (a requirement of some methods tested in this study), only 1 observation for each cow was retained for analysis. Where multiple records existed for an animal, the Mahalanobis distances between the average principal component (**PC**) scores from the entire data set and the multiple observations from the animal were computed. The observation with the greatest distance was retained, with the aim of maximizing the variability in the data set.

High-noise-level regions (Hewavitharana and van Brakel, 1997) were removed from each spectrum; the spectral regions between 1,710 and 1,600 cm^{-1}, between 3,690 and 2,990 cm^{-1}, and $>$3,822 cm^{-1} were discarded. Consequently, a total of 531 wavelengths were used for the analyses. The transmittance values of the wavelengths were transformed to absorbance values by taking the $\log_{10}$ of the reciprocal of the transmittance value. Outliers for the traits of interest were defined as those $>$3 standard deviations (SD) from the mean of the respective trait and were subsequently removed from the analysis of that trait. The 16 noncoagulating milk samples were removed from the analyses of RCT, k20, a30, and a60. The numbers of samples as well as the mean, SD, median, minimum and maximum, coefficient of variation, and skewness of all traits investigated are provided in Table 1.

To assess the utility of classification-based statistical methods, the milk technological traits were divided into 2 classes based on their respective median value; the median was chosen as a threshold to split these traits into high and low, representing a proxy for the suitability of milk for cheese production. The content of each protein was divided into 4 classes based on quartiles,

with the aim of reducing the range in values within each class. In a separate series of analyses, noncoagulating samples (n = 16) were rejoined to the data set to test the ability of classification models to discriminate between coagulating and noncoagulating samples.

Statistical Analyses

To compare the performance of the different statistical ML approaches, the data were divided into 4 subdata sets with approximately the same numbers of observations in each, and 4-fold cross-validation was performed. The data division and cross-validation were performed separately for each trait, using the fold function in the groupdata2 package (Olsen, 2020) in R (R Core Team, 2020) to balance data across folds. All the analyses were conducted using the statistical software R 3.6.1.

Regression-Based Approaches

Eleven different regression-based statistical ML methods were explored; James et al. (2017) provide an excellent review of such methods. For some of the approaches, the tuning parameters were user defined or selected via cross-validation, and for others the default settings were used.

Partial least squares regression is a supervised dimension reduction method (Geladi and Kowalski, 1986). Partial least squares regression seeks out a small number of new variables (i.e., factors) that are linear combinations of the wavelengths, exploiting information on the response variable in doing so. Thus, PLSR uses both the trait data and the spectra to detect directions in the data space that best explain both. Partial least squares regression then fits a linear regression model via least squares to the trait data and the generated factors. Because a large portion of the information in the original data is captured by the generated factors, and because they are fewer in number, overfitting is mitigated. The number of PLSR factors to generate is data-dependent and user defined, typically by examining the change in root mean square error (**RMSE**) with each additional factor. In the present study, leave-one-out cross-validation was used to choose the number of factors to use in the model. A different number of factors were used in each of the 4 folds. The R package pls (Mevik et al., 2019) was used here to implement PLSR.

Principal component regression is similar in nature to PLSR but instead uses a linear regression model (estimated via least squares) of the trait on a small number of PC derived from the spectra alone. Similar to PLSR, a small number (compared with the number of wavelengths) of PC generally suffice to explain most of the variability in the data. The number of PC to retain is user defined, here by examining the change in RMSE with each additional PC. The R package pls (Mevik et al., 2019) was again used to implement PC regression.

Projection pursuit regression (**PPR**; Friedman and Stuetzle, 1981) is similar to both PLSR and PC regression in that it extracts linear combinations of the wavelengths as new derived features. Projection pursuit regression then models the trait as a nonlinear function of the newly derived features, where the prediction process uses flexible smoothing methods. The R package stats (R Core Team, 2020) was used to apply PPR.

Ridge regression (**RR**; Hoerl and Kennard, 1970) fits a linear regression model that includes all wavelengths

Table 1. Number of samples (No.), mean, SD, median, minimum (Min) and maximum (Max), CV, and skewness for the technological traits and protein traits considered

Trait	No.	Mean	SD	Median	Min	Max	CV	Skewness
Technological[1]								
RCT, min	482	20.81	9.02	19.50	1.75	47.50	0.43	0.53
k20, min	439	5.82	3.42	5.00	1.25	15.75	0.59	0.98
a30, mm	401	32.24	15.62	31.48	2.02	74.90	0.48	0.18
a60, mm	478	31.28	11.73	29.32	1.76	66.34	0.38	0.69
CN micelle size, mm	553	172.92	26.02	168.70	109.10	250.30	0.15	0.65
pH	601	6.69	0.10	6.68	6.41	6.97	0.02	0.23
Heat stability, min	431	9.43	7.22	6.80	0.58	31.00	0.77	1.39
Protein, g/L								
α_{S1}-CN	447	14.09	2.39	13.98	7.21	20.86	0.17	0.23
α_{S2}-CN	460	3.67	0.94	3.60	0.88	6.36	0.26	0.14
β-CN	449	12.80	2.16	12.64	6.39	19.09	0.17	0.20
κ-CN	453	5.77	1.43	5.71	1.55	9.54	0.25	−0.03
α-LA	457	1.12	0.30	1.09	0.23	2.00	0.27	0.42
β-LG A	462	2.49	1.17	2.31	0.36	5.90	0.47	0.53
β-LG B	464	2.45	1.68	2.47	0.00	7.44	0.69	0.41

[1]RCT = rennet coagulation time; k20 = curd-firming time; a30, a60 = curd firmness at 30 and 60 min, respectively.

but shrinks each regression coefficient estimated separately toward zero during model fitting. This approach, known as regularization, reduces the variance of predictions, at the expense of an increase in their bias. Although RR is not a dimension reduction method and includes all wavelengths, it is computationally efficient as it fits only a single model. User specification of a tuning parameter is required, which here was selected by cross-validation. The package glmnet (Friedman et al., 2010) was used for the RR analysis in this study.

Although RR shrinks the regression coefficients toward zero, it does not shrink any to exactly zero (except when the tuning parameter is infinite) and so all variables are always included in the prediction model. This can result in good prediction accuracy but poor model interpretability. Least absolute shrinkage and selection operator (**LASSO**; Tibshirani, 1996) is similar in nature to RR but allows coefficient estimates to be exactly zero and, hence, is also a variable selection method that results in more interpretable models. The tuning parameter was selected based on the lowest mean cross-validated error. The R package glmnet was again used for the LASSO analysis.

Elastic net (**EN**; Zou and Hastie, 2005) offers a compromise between RR and the LASSO, in that it selects wavelengths similar to the LASSO but shrinks the coefficients of correlated wavelengths together like RR. Thus, EN can be considered a dimension reduction method, although it will select more wavelengths than the LASSO. The R package used for the EN analyses was glmnet.

Model averaging (**MA**) is a novel approach that consists of averaging the predictions from several of the previously considered approaches, which, in the present study, were PLSR, RR, LASSO, and EN. These models were selected due to their similarity in approach.

Spike and slab regression (**SSR**; Mitchell and Beauchamp, 1988) takes a Bayesian approach by assuming a bimodal prior distribution for the regression coefficients, with one mode at zero and one nonzero mode, followed by the use of a generalized EN to fit the model. The R package used for the analyses was spikeslab (Ishwaran et al., 2010).

Random forests (**RF**) produce multiple decision trees (**DT**), the predictions from which are combined to give a consensus prediction. Decision tree–based methods (Breiman et al., 1984) are so called because they can be summarized visually by a tree-like structure. Decision trees work by segmenting the predictor space into several simple regions. Prediction for a test spectrum is simply the mean of the training observations in the region to which the test spectrum belongs. The predictor space is segmented recursively, beginning with a root node and subsequently creating branches determined by splitting rules based on the predictor values. The terminal nodes or leaves of the resulting tree define the simple regions used for prediction. However, DT suffer from high variance in its response, which RF overcomes by averaging predictions from many DT, but where at each split only a random sample of wavelengths are considered. The number of DT and the number of wavelengths randomly sampled as candidates at each split is user defined. Here 500 DT were used, and the number of wavelengths considered at each split was the number of wavelengths divided by 3. The R package used for the analyses was randomForest (Liaw and Wiener, 2002).

Boosting is a general concept that can be used with many statistical ML methods to improve predictions. Unlike the RF setting, each DT is fitted to a modified version of the original data set. The trees are grown sequentially, where, at each stage, a tree is fitted to the residuals from the previous model fit, thus improving model fit in areas of the predictor space where performance in a single DT was poor. Boosting requires the specification of several settings: here the number of trees considered was 500, and the shrinkage parameter was set to 0.01. The approach was implemented using the gbm R package (Greenwell et al., 2019).

Neural networks (**NN**) are nonlinear generalizations of a linear model. In a regression setting, NN first construct derived features that are linear combinations of the wavelengths. The outcome variable is then modeled as a function of linear combinations of the derived features. An NN is typically represented in a network diagram with several hidden layers, each representing different functions of the derived features. Here, a 2-layer NN was fitted, with Bayesian regularization employed to improve generalizability, using the R package brnn (Perez Rodriguez and Gianola, 2020).

Classification Approaches

The outcome traits were divided into classes and the performance of 4 classification approaches assessed.

Partial least square discriminant analysis (**PLSDA**) is a dimension reduction model used for classification purposes. Partial least square discriminant analysis works similarly to PLSR, but for the former the response variable is dichotomized. The model proceeds similarly to PLSR, with prediction to the classes determined by whether or not the output is greater than a specified threshold. The R package used for the analysis was caret (Kuhn, 2020).

Random forests applied for classification purposes follows the procedure previously described for RF for

regression purposes. The implementation of RF for classification purposes used the same number of trees as in the regression setting, but the number of wavelengths considered at each split was set to the square root of the number of wavelengths.

Boosting DT were also used for classification purposes, with the number of trees considered remaining at 500, as in the regression setting.

Support vector machine (**SVM**) is a classification method that allows for nonlinear decision boundaries between classes by enlarging the feature space using kernels. In the enlarged space, the boundary is linear, but in the wavelength space, the boundary is nonlinear and more flexible. The R package used for the SVM analyses was e1071 (Meyer et al., 2019), in which a linear kernel was employed.

Measures of Prediction Performance

The performance of each regression method was evaluated by examining the RMSE from the calibration data (3 folds of the data), the root mean square error from the cross-validation data (the remaining fold; **RMSEV**), and the coefficient of determination (R^2, from both the calibration and the cross-validation data). Furthermore, the slope coefficient of a simple linear regression of the observed on the predicted value of each trait, as well as the bias corresponding to the mean of the observed minus the mean of the predicted values of the trait were obtained from the cross-validation data. The ratio of performance to interquartile distance (**RPIQ**) was used to assess the model consistency (Bellon-Maurel et al., 2010). The RPIQ is calculated as the ratio between the interquartile range of the observed trait values and the RMSE. The RPIQ was used in the present study instead of ratio performance deviation because it is better suited to non-normally distributed traits. Given a lack of evidence to support the use of threshold values for interpretation (Bellon-Maurel et al., 2010), the RPIQ was used in this study to compare performance of alternative models rather than to quantify prediction accuracy of specific traits per se.

The performance of each classification method for the milk technological traits was evaluated by examining the area under the receiver operating characteristic curve, the sensitivity (i.e., proportion of the high class correctly classified), the specificity (i.e., proportion of the low class correctly classified), and the accuracy (i.e., the ratio of the number of correctly classified observations to the total number of observations). For the milk proteins, which were divided into 4 classes, the classification methods' performance was assessed by the accuracy (i.e., the ratio of the number of correctly

classified observations to the total number of observations).

In the regression analyses, the RMSE, R^2, RMSEV, bias, and RPIQ were calculated as the average of the 4 folds of calibration or cross-validation data. The standard deviation (SD) of RMSE, R^2, RMSEV, bias, and RPIQ across folds were also calculated, thus reflecting the variability or robustness across folds. The slope and its standard error in the regression analyses were estimated once across the entire data set of predicted values (i.e., across all 4 folds). The prediction performance for classification was calculated as the average of the 4 folds of calibration or cross-validation data; the SD reflects the variability across folds. In the present study, when a continuous trait was investigated, the RMSEV was used to identify the "best" model. When a trait in question was a categorical trait, the accuracy was used to identify the "best" model.

RESULTS

Supplemental Tables S1 to S18 and Supplemental Figures S1 to S3 are available on the Teagasc Open Access Repository, T-Stór (https://hdl.handle.net/11019/2390; Frizzarin et al., 2021).

Prediction of Continuous Traits

Table 2 details the regression model with the lowest RMSEV for each trait, the RMSEV obtained, and the coefficient of determination in the cross-validation data set. The difference between the RMSEV of the "best" prediction model and the corresponding RMSEV obtained from PLSR on the same trait is also detailed. The MA approach most frequently performed "best" across all traits, having the lowest RMSEV for 6 of the 14 traits investigated (i.e., a60, α_{S1}-CN, α_{S2}-CN, κ-CN, α-LA, and β-LG B). The LASSO and NN methods performed similarly to MA for κ-CN prediction. The NN had the lowest RMSEV for all of RCT, k20, and heat stability prediction, whereas LASSO had the lowest RMSEV for CMS prediction. Ridge regression had the lowest RMSEV for a30, β-CN, and β-LG A, but PLSR had the lowest RMSEV for pH. The average difference in RMSEV between PLSR and the "best" model varied from 0.18% (κ-CN) to 3.67% (heat stability). The prediction performance for each of the milk technological traits is presented in Supplemental Tables S1 to S7 (https://hdl.handle.net/11019/2390). Supplemental Tables S8 to S14 (https://hdl.handle.net/11019/2390) summarize the prediction performance of the different models for all the milk proteins.

The number of factors selected by PLSR across folds was consistent for some traits (± 2 factors), although

Table 2. Summary of the regression method with the lowest root mean square error in cross-validation (RMSEV) and the percentage difference in RMSEV between the "best" algorithm and PLSR for each trait investigated with the respective prediction performance and SD across folds

Trait	Best regression method[1]	RMSEV (SD)	R^2 (SD)	Comparison to PLSR (% difference)
Technological trait[2]				
RCT, min	NN	6.397 (0.692)	0.50 (0.03)	2.36
k20, min	NN	2.770 (0.280)	0.36 (0.06)	1.61
a30, mm	RR	12.495 (0.084)	0.37 (0.05)	1.80
a60, mm	Average	10.245 (0.972)	0.25 (0.10)	0.59
CN micelle size, mm	LASSO	25.286 (1.294)	0.08 (0.03)	1.42
pH	PLSR	0.061 (0.002)	0.65 (0.04)	—
Heat stability, min	NN	5.464 (0.390)	0.45 (0.05)	3.67
Protein, g/L				
α_{S1}-CN	Average	1.745 (0.136)	0.47 (0.05)	2.55
α_{S2}-CN	Average	0.734 (0.113)	0.38 (0.04)	0.54
β-CN	RR	1.759 (0.128)	0.35 (0.05)	2.30
κ-CN	LASSO, Average, NN	1.095 (0.027, 0.027, 0.027)	0.42 (0.09, 0.09, 0.09)	0.18
α-LA	Average	0.255 (0.020)	0.27 (0.02)	0.39
β-LG A	RR	1.050 (0.113)	0.19 (0.04)	0.66
β-LG B	Average	1.443 (0.117)	0.27 (0.08)	1.24

[1]PLSR = partial least square regression; RR = ridge regression; EN = elastic net; Average = model averaging approach; NN = neural network; LASSO = least absolute shrinkage and selection operator.

[2]RCT = rennet coagulation time; k20 = curd-firming time; a30, a60 = curd firmness at 30 and 60 min, respectively.

for others the number of factors selected varied across folds (±8 factors). Notably, the number of wavelengths selected for use in the model varied according to trait and model; SSR selected on average between 0.25 (κ-CN) and 93.5 (RCT) fewer wavelengths than LASSO, whereas EN always selected more wavelengths than either LASSO or SSR. The numbers of wavelengths selected by LASSO, EN, and SSR for each trait are presented in Supplemental Table S15 (https://hdl.handle.net/11019/2390), and the subsets of wavelengths selected are presented graphically in Supplemental Figures S1 to S3 (https://hdl.handle.net/11019/2390). The different models tended to select similar subsets of wavelengths. Also, PLSR, RR, and RF attributed greatest coefficients to these regions. In particular, the regions between 1,100 and 1,000 cm^{-1}, between 1,530 and 1,462 cm^{-1}, between 1,790 and 1,735 cm^{-1}, and between 3,730 and 3,710 cm^{-1} were important for several traits. The region between 1,100 and 1,000 cm^{-1} was recurrently present for all the investigated traits, with the exception of β-CN. The region between 1,530 and 1,430 cm^{-1} was present for all the protein traits, as well as for RCT, a60, CMS, and pH. The region between 1,790 and 1,735 cm^{-1} was present for all the milk technological traits with the exception of a30 and was present for α-LA and β-LG A. The region between 3,730 and 3,710 cm^{-1} was present for a30, a60, pH, α_{S1}-CN, β-CN, κ-CN, and β-LG B. In this specific region, for a60, pH, α_{S2}-CN, and β-LG B, the wavelength 3,726 cm^{-1} was always selected; for α_{S1}-CN, β-CN, and κ-CN, the wavelength 3,714 cm^{-1} was always selected.

Prediction of Categorical Traits

Table 3 summarizes the "best" prediction model and its prediction accuracy across all traits. Support vector machine was the method with the greatest accuracy for 6 of the 7 binary technological traits investigated (i.e., RCT, k20, a30, CMS, pH, and heat stability); PLSDA had the same accuracy as SVM for RCT, pH, and heat stability prediction. Partial least squares discriminant analysis was the model with the greatest accuracy also for a60 prediction. For the binary technological traits, the greatest average accuracy was for pH prediction (0.80, SD = 0.03, 0.02), and the lowest average accuracy was for CMS (0.62, SD = 0.03). Sensitivity of discrimination of coagulating samples ranged from 0.98 (SD = 0.02; boosting DT) to 1.00 (SD = 0.00; PLSDA), but specificity was poor and ranged from 0.44 (PLSDA, RF, SVM) to 0.50 (boosting DT).

When the protein traits were split into quartiles for prediction, PLSDA had the greatest accuracy for 3 of the 6 traits (i.e., α_{S2}-CN, β-LG A, and β-LG B). Support vector machine produced the greatest accuracy for 2 traits (i.e., β-CN, and α-LA), and RF had the greatest accuracy for the remaining 2 traits (i.e., α_{S1}-CN and κ-CN). When the protein traits were divided into quartiles, accuracy ranged from 0.40 (SD = 0.04; α_{S2}-CN) to 0.48 (SD = 0.02; α_{S1}-CN). The prediction performance for the milk technological traits when classified are shown in Supplemental Table S16 (https://hdl.handle.net/11019/2390). Supplemental Tables S17 and S18 (https://hdl.handle.net/11019/2390) summa-

rize the prediction performances for classified CN and whey proteins, respectively.

DISCUSSION

Although traditional statistical methods have served the prediction of phenotypes from milk spectral data well for several cattle (McParland and Berry, 2016) and milk traits (De Marchi et al., 2014), the objective of the present study was to evaluate alternative statistical approaches with a focus on ML techniques.

Prediction of Continuous Traits

Partial least squares regression is considered the benchmark method, given its consistently strong prediction performance in chemometric analyses (Wold et al., 2001). However, PLSR did not consistently perform the best for the traits considered in the present study. With the exception of pH, the average difference in RMSEV between PLSR and the best prediction method ranged from 0.18% (κ-CN) to 3.67% (heat stability). Nonetheless, although variable prediction accuracy was observed across cross-validation folds, the "best" overall method was generally consistently the best in each fold. For example, when heat stability was predicted, NN always outperformed PLSR, with the RMSEV ranging from 0.76% to 6.03% lower with the NN method. Other methods investigated here, such as PPR, performed poorly, possibly due to the difficulty in

Table 3. Summary of the classification method with the greatest accuracy for each trait investigated with the respective prediction performance and SD across folds

Trait	Best classification	
	Method[1]	Accuracy[2] (SD)
Technological[3]		
RCT	PLSDA, SVM	0.75 (0.03, 0.06)
k20	SVM	0.73 (0.02)
a30	SVM	0.73 (0.03)
a60	PLSDA	0.69 (0.07)
CN micelle size	SVM	0.62 (0.03)
pH	PLSDA, SVM	0.80 (0.03, 0.02)
Heat stability	PLSDA, SVM	0.74 (0.04, 0.05)
Protein		
α_{S1}-CN	RF	0.48 (0.02)
α_{S2}-CN	PLSDA	0.40 (0.04)
β-CN	SVM	0.46 (0.04)
κ-CN	RF	0.45 (0.02)
α-LA	SVM	0.43 (0.03)
β-LG A	PLSDA	0.42 (0.03)
β-LG B	PLSDA	0.41 (0.04)

[1]PLSDA = partial least squares discriminant analysis; RF = random forest; SVM = support vector machine.

[2]Proportion of correctly classified observations.

[3]RCT = rennet coagulation time; k20 = curd-firming time; a30, a60 = curd firmness at 30 and 60 min, respectively.

choosing the correct tuning parameters, which requires careful specification of many settings by the user. Thus, the examined methods demonstrated better or comparable performance to the traditionally used PLSR, in line with Wolpert and Macready's (1997) assertion that for any algorithm, any superior performance in one class of problems is offset by its performance in another class. The "best" model varied depending on the data distribution and the range and variability present in the trait under investigation. Therefore, the same trait in a different data set could potentially be "best" predicted using a different method, and practitioners should consider these methods when predicting milk traits from mid-infrared data. Other examples of methods compared in the literature for predictive performance also gave inconsistent results across studies (e.g., Ferrand-Calmels et al., 2014; Bonfatti et al., 2017; El Jabri et al., 2019). The variability in performance is likely related to differences in the traits predicted and data sets used. Notwithstanding this, the MA approach draws strength from averaging across several methods, resulting in more accurate predictions than those achieved by any of the individual methods alone.

Shrinkage methods are used in genomic prediction (Li and Sillanpää, 2012; Ogutu et al., 2012; Azevedo et al., 2015) for dimension reduction. Indeed, shrinkage methods should identify the variables most strongly related to the trait being predicted. Hence, in chemometric analyses, it is expected that shrinkage methods could also be used to identify the most informative wavelengths where wavelengths may be considered to be analogous to SNPs in genomic predictions. However, the literature presents contrasting results about the potential of shrinkage methods (e.g., Bonfatti et al., 2017; El Jabri et al., 2019). Other methodologies not based on shrinkage models have been developed for wavelength selection in spectroscopy (Gottardo et al., 2015; Vohland et al., 2014). In the present study, 3 variable selection approaches were investigated, namely, LASSO, EN, and SSR. These 3 methods shrink to zero the coefficients of the wavelengths not deemed to be related to the trait under investigation. All remaining methods considered all the wavelengths available in the data set for the prediction, giving different weights (coefficients) to each wavelength but never attributing zero as a coefficient. In the present study, LASSO was the "best" model for 2 of the 14 traits investigated. Combining information from (1) a cross-investigation of the wavelengths selected by wavelength selection models (LASSO, EN, and SSR), (2) the coefficients calculated by PLSR and RR, and (3) the variable importance in RF, can inform which wavelength regions are related to specific traits. In fact, LASSO, EN, and SSR partly selected the same wavelengths, and these wavelengths

were also associated with the greatest coefficients in both PLSR and RR as well as the greatest importance in RF; thus, specific wavelengths were identified as important for trait prediction across models. Requiring only selected wavelengths in the prediction model of milk constituents could justify the development of an instrument focused solely on these specific wavelength regions to predict prespecified groups of traits, for example, milk coagulation traits or milk proteins. Such an instrument should have reduced construction (and thus purchase) costs, making it more amenable for more widespread in-line use.

The data set used in the current study was a subset of that previously used to quantify the potential of MIRS as a predictor of individual milk proteins (McDermott et al., 2016) and technological traits (Visentin et al., 2015) using PLSR. Different editing of the original data set was required in the current study to satisfy the assumptions of some of the ML methods employed here. Further, the handling of the data was different in the current study, where the data set was divided into 4 subdata sets or folds (25% in each fold) to perform 4-fold cross-validation. Visentin et al. (2015) randomly divided the data set once into calibration and validation data sets, with 80% of the data included in the calibration data set. Thus, the PLSR results reported in the current study are not identical to those in previous studies.

Prediction of Categorized Traits

Some traits, including RCT, k20, a30, a60, and pH, can be used together to define milk suitability for cheese making. Manuelian et al. (2017) investigated the ability of PLSR applied to MIRS to predict milk coagulation traits in Mediterranean buffalo; Manuelian et al. (2017) reported a coefficient of determination in cross-validation varying from 0.27 for k20 prediction to 0.76 for pH prediction. After this, Manuelian et al. (2017) categorized the samples into noncoagulating milk and coagulating milk with the purpose of discriminating the samples based on their milk coagulating ability; the model correctly classified 91.57% and 67.86% of noncoagulating milk samples in the calibration and validation sets, respectively. Results from the present study reveal a poor discrimination between coagulating and noncoagulating milk, likely due to the unbalanced data available; only 3.2% of samples were considered noncoagulating in the data set used here.

Although different studies reported the potential of predicting classes, either by clustering similar traits or by dividing a specific trait in classes (Manuelian et al., 2017; Grelet et al., 2019; Duplessis et al., 2020), accurate methods that permit the comparison of regression results and classification results are needed to enable appropriate conclusions about the optimal approach; unfortunately, no such statistical method currently exists. Which approach should be used depends on the context and on the type of variable to be predicted. These studies used canonical discriminant analysis (Manuelian et al., 2017) or PLSDA (Grelet et al., 2019; Duplessis et al., 2020) to perform class prediction, but the SVM has been shown in the present study to be a possible alternative due most likely to its ability to exploit nonlinear associations between the wavelengths and the trait.

Practical Utility

Milk coagulation properties such as greater curd-firming capacity and shorter milk coagulation time are correlated with improved sensory properties of cheese as well as with greater cheese yield (Martin et al., 1997; Pretto et al., 2013). Heat stability, CMS, and pH are fundamental traits for cheese production and production of other milk-related products such as milk powder (Singh, 2004). Similarly, α_{S1}-CN, β-CN, κ-CN, and β-LG B in milk have positive effects on cheese yield (Wedholm et al., 2006). Although the ML methods investigated here only slightly improved predictions over PLSR, and only for some traits, and despite the prediction accuracy remaining low, the modern ML methods investigated in the present study clearly demonstrate promise. Improved prediction of traits, however small, is useful for the milk processing industry to discriminate milk at the preprocessing stage, enabling milk to be used for the process for which it is most suited.

Although some of the improvements in accuracy may be small, they can be generated at no additional physical or computational expense; the run times for the methods considered are of a similar order of magnitude to run times for PLS approaches. Further, all the methods considered can be easily implemented on a standard personal computer. As such, the modern statistical ML methods have similar practical utility to currently used methods.

CONCLUSIONS

In chemometric analyses, more interpretable methodologies (e.g., PLSR or LASSO) should be preferred to more complicated methods (e.g., NN) if the results do not differ, due to their interpretability and their reliance on fewer tuning parameters. In the current study, the "best" method varied depending on the data distribution and the range and variability present in the trait under investigation. Several methods demonstrated better or comparable performance to the traditionally used PLS-

based methods, and practitioners should consider such alternative methods when predicting milk traits from mid-infrared data. The MA approach was the method that most often had the lowest RMSEV, and its use and implementation should be considered for regression analyses. The division of continuous traits into classes can be a useful solution for traits poorly predicted with regression methods. However, prediction of traits that were divided into more than 2 classes performed poorly here. Although accuracy of prediction of traits in this study was moderate, the application of novel statistical ML methods may improve the prediction of milk traits; however, the well-established PLSR-based method still performed competitively. (Code generated for use in this study is available at https://github.com/maria -ire/Code-used-for-milk-quality-traits-predicted-from -routinely-available-milk-spectra-Paper-analyses.)

REFERENCES

Azevedo, C. F., M. D. V. de Resende, F. F. E Silva, J. M. S. Viana, M. S. F. Valente, M. F. R. Resende Jr., and P. Muñoz. 2015. Ridge, LASSO and Bayesian additive-dominance genomic models. BMC Genet. 16:105. https://doi.org/10.1186/s12863-015-0264-2.

Bellon-Maurel, V., E. Fernandez-Ahumada, B. Palagos, J. M. Roger, and A. McBratney. 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. Trends Analyt. Chem. 29:1073–1081. https://doi.org/10.1016/j.trac.2010.05.006.

Bonfatti, V., F. Tiezzi, F. Miglior, and P. Carnier. 2017. Comparison of Bayesian regression models and partial least squares regression for the development of infrared prediction equations. J. Dairy Sci. 100:7306–7319. https://doi.org/10.3168/jds.2016-12203.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. Classification and Regression Trees. Wadsworth.

Cecchinato, A., M. De Marchi, L. Gallo, G. Bittante, and P. Carnier. 2009. Mid-infrared spectroscopy predictions as indicator traits in breeding programs for enhanced coagulation properties of milk. J. Dairy Sci. 92:5304–5313. https://doi.org/10.3168/jds.2009-2246.

Davies, D. T., and J. C. D. White. 1966. The stability of milk protein to heat: I. Subjective measurement of heat stability of milk. J. Dairy Res. 33:67–81. https://doi.org/10.1017/S0022029900011730.

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.

Dehareng, F., C. Delfosse, E. Froidmont, H. Soyeurt, C. Martin, N. Gengler, A. Vanlierde, and P. Dardenne. 2012. Potential use of milk mid-infrared spectra to predict individual methane emission of dairy cows. Animal 6:1694–1701. https://doi.org/10.1017/S1751731112000456.

Duplessis, M., D. Pellerin, C. L. Girard, D. E. Santschi, and H. Soyeurt. 2020. Short communication: Potential prediction of vitamin B_{12} concentration based on mid-infrared spectral data using Holstein Dairy Herd Improvement milk samples. J. Dairy Sci. 103:7540–7546. https://doi.org/10.3168/jds.2019-17758.

El Jabri, M., M. P. Sanchez, P. Trossat, C. Laithier, V. Wolf, P. Grosperrin, E. Beuvier, O. Rolet-Répécaud, S. Gavoye, Y. Gaüzère, O. Belysheva, E. Notz, D. Boichard, and A. Delacroix-Buchet. 2019. Comparison of Bayesian and partial least squares regression methods for mid-infrared prediction of cheese-making properties in Montbéliarde cows. J. Dairy Sci. 102:6943–6958. https://doi.org/10.3168/jds.2019-16320.

Ferrand-Calmels, M., I. Palhière, M. Brochard, O. Leray, J.-M. Astruc, M.-R. Aurel, S. Barbey, F. Bouvier, P. Brunschwig, H. Caillat, M. Douguet, F. Faucon-Lahalle, M. Gelé, G. Thomas, J. M. Trommenschlager, and H. Larroque. 2014. Prediction of fatty acid profiles in cow, ewe, and goat milk by mid-infrared spectrometry. J. Dairy Sci. 97:17–35. https://doi.org/10.3168/jds.2013-6648.

Friedman, J., T. Hastie, and R. Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33:1–22. https://doi.org/10.18637/jss.v033.i01.

Friedman, J. H., and W. Stuetzle. 1981. Projection pursuit regression. J. Am. Stat. Assoc. 76:817–823. https://doi.org/10.1080/01621459.1981.10477729.

Frizzarin, M., I. C. Gormley, D. Berry, T. B. Murphy, A. Casa, A. Lynch, and S. McParland. 2021. Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. http://hdl.handle.net/11019/2390.

Geladi, P., and B. R. Kowalski. 1986. Partial least-squares regression: A tutorial. Anal. Chim. Acta 185:1–17. https://doi.org/10.1016/0003-2670(86)80028-9.

Gottardo, P., M. De Marchi, M. Cassandro, and M. Penasa. 2015. Technical note: Improving the accuracy of mid-infrared prediction models by selecting the most informative wavelengths. J. Dairy Sci. 98:4168–4173. https://doi.org/10.3168/jds.2014-8752.

Greenwell, B., B. Boehmke, J. Cunningham, and G. B. M. Developers. 2019. gbm: Generalized Boosted Regression Models. R package version 2.1.5. https://CRAN.R-project.org/package=gbm.

Grelet, C., A. Vanlierde, M. Hostens, L. Foldager, M. Salavati, K. L. 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.

Hewavitharana, A. K., and B. van Brakel. 1997. Fourier transform infrared spectrometric method for the rapid determination of casein in raw milk. Analyst (Lond.) 122:701–704. https://doi.org/10.1039/a700953d.

Hoerl, A. E., and R. W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12:55–67. https://doi.org/10.1080/00401706.1970.10488634.

Ishwaran H., U. B. Kogalur and J. S. Rao. 2010. Spikeslab: Prediction and variable selection using spike and slab regression. R package version 1.1.2. https://cran.r-project.org/web/packages/spikeslab/spikeslab.pdf. Accessed Mar. 31, 2021.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2017. An Introduction to Statistical Learning with Applications in R. Springer.

Keller, L. P., S. Bajt, G. A. Baratta, J. Borg, J. P. Bradley, D. E. Brownlee, H. Busemann, J. R. Brucato, M. Burchell, L. Colangeli, L. d'Hendecourt, Z. Djouadi, G. Ferrini, G. Flynn, I. A. Franchi, M. Fries, M. M. Grady, G. A. Graham, F. Grossemy, A. Kearsley,

G. Matrajt, K. Nakamura-Messenger, V. Mennella, L. Nittler, M. E. Palumbo, F. J. Stadermann, P. Tsou, A. Rotundi, S. A. Sandford, C. Snead, A. Steele, D. Wooden, and M. Zolensky. 2006. Infrared spectroscopy of comet 81P/Wild 2 samples returned by Stardust. Science 314:1728–1731. https://doi.org/10.1126/science.1135796.

Kuhn, M. 2020. caret: Classification and Regression Training. R package version 6.0-85. https://cran.r-project.org/web/packages/caret/index.html. Accessed Mar. 31, 2021.

Li, B., N. Zhang, Y.-G. Wang, A. W. George, A. Reverter, and Y. Li. 2018. Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. Front. Genet. 9:237. https://doi.org/10.3389/fgene.2018.00237.

Li, Z., and M. J. Sillanpää. 2012. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. Theor. Appl. Genet. 125:419–435. https://doi.org/10.1007/s00122-012-1892-9.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2:18–22.

Manuelian, C. L., G. Visentin, C. Boselli, G. Giangolini, M. Cassandro, and M. De Marchi. 2017. Short communication: Prediction of milk coagulation and acidity traits in Mediterranean buffalo milk using Fourier-transform mid-infrared spectroscopy. J. Dairy Sci. 100:7083–7087. https://doi.org/10.3168/jds.2017-12707.

Martin, B., J.-F. Chamba, J.-B. Coulon, and E. Perreard. 1997. Effect of milk chemical composition and clotting characteristics on chemical and sensory properties of Reblochon de Savoie cheese. J. Dairy Res. 64:157–162. https://doi.org/10.1017/S0022029996001975.

McDermott, A., G. Visentin, M. De Marchi, D. Berry, M. Fenelon, P. O'connor, O. Kenny, and S. McParland. 2016. Prediction of individual milk proteins including free amino acids in bovine milk using mid-infrared spectroscopy and their correlations with milk processing characteristics. J. Dairy Sci. 99:3171–3182. https://doi.org/10.3168/jds.2015-9747.

McParland, S., and D. P. Berry. 2016. The potential of Fourier transform infrared spectroscopy of milk samples to predict energy intake and efficiency in dairy cows. J. Dairy Sci. 99:4056–4070. https://doi.org/10.3168/jds.2015-10051.

McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. Pryce, and D. P. Berry. 2014. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. J. Dairy Sci. 97:5863–5871. https://doi.org/10.3168/jds.2014-8214.

Mevik, B. H., R. Wehrens, and K. H. Liland. 2019. pls: Partial Least Squares and Principal Component Regression. R package version 2.7-2. Accessed Nov. 10, 2020. https://CRAN.R-project.org/package=pls.

Meyer, D., E. Dimitriadou. K. Hornik, A. Weingessel, F. Leisch, C. Chang, and C. Lin. 2019. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.7-3. Accessed Nov. 10, 2020. https://CRAN.R-project.org/package=e1071.

Mitchell, T. J., and J. J. Beauchamp. 1988. Bayesian variable selection in linear regression. J. Am. Stat. Assoc. 83:1023–1032. https://doi.org/10.1080/01621459.1988.10478694.

Ogutu, J. O., T. Schulz-Streeck, and H.-P. Piepho. 2012. Genomic selection using regularized linear regression models: ridge regression, LASSO, elastic net and their extensions. Proc. 15th European Workshop on QTL Mapping and Marker Assisted Selection (QTLMAS). BMC Proceedings 6:S10. https://doi.org/10.1186/1753-6561-6-S2-S10.

Olsen, L. R. 2020. groupdata2: Creating Groups from Data. R package version 1.2.0. Accessed Nov. 10, 2020. https://CRAN.R-project.org/package=groupdata2.

Perez Rodriguez, P., and D. Gianola. 2020. brnn: Bayesian Regularization for Feed-Forward Neural Networks. R package version 0.8. Accessed Nov. 10, 2020. https://CRAN.R-project.org/package=brnn.

Petrich, W. 2001. Mid-infrared and Raman spectroscopy for medical diagnostics. Appl. Spectrosc. Rev. 36:181–237. https://doi.org/10.1081/ASR-100106156.

Pretto, D., M. De Marchi, M. Penasa, and M. Cassandro. 2013. Effect of milk composition and coagulation traits on Grana Padano cheese yield under field conditions. J. Dairy Res. 80:1–5. https://doi.org/10.1017/S0022029912000453.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Accessed Nov. 10, 2020. https://www.R-project.org/.

Singh, H. 2004. Heat stability of milk. Int. J. Dairy Technol. 57:111–119. https://doi.org/10.1111/j.1471-0307.2004.00143.x.

Skolik, P., M. R. McAinsh, and F. L. Martin. 2018. Biospectroscopy for plant and crop science. Pages 15–49 in Comprehensive Analytical Chemistry. J. Lopes and C. Sousa C, ed. Elsevier. https://doi.org/10.1016/bs.coac.2018.03.001.

Soyeurt, H., P. Dardenne, F. Dehareng, G. Lognay, D. Veselko, M. Marlier, C. Bertozzi, P. Mayeres, and N. Gengler. 2006. Estimating fatty acid content in cow milk using mid-infrared spectrometry. J. Dairy Sci. 89:3690–3695. https://doi.org/10.3168/jds.S0022-0302(06)72409-2.

Tibshirani, R. 1996. Regression shrinkage and selection via the LASSO. J. R. Stat. Soc. B 58:267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Visentin, G., A. McDermott, S. McParland, D. P. Berry, O. Kenny, A. Brodkorb, M. A. Fenelon, and M. De Marchi. 2015. Prediction of bovine milk technological traits from mid-infrared spectroscopy analysis in dairy cows. J. Dairy Sci. 98:6620–6629. https://doi.org/10.3168/jds.2015-9323.

Visentin, G., M. Penasa, P. Gottardo, M. Cassandro, and M. De Marchi. 2016. Predictive ability of mid-infrared spectroscopy for major mineral composition and coagulation traits of bovine milk by using the uninformative variable selection algorithm. J. Dairy Sci. 99:8137–8145. https://doi.org/10.3168/jds.2016-11053.

Visser, S., C. J. Slangen, and H. S. Rollema. 1991. Phenotyping of bovine milk proteins by reversed-phase high performance liquid chromatography. J. Chromatogr. 548:361–370. https://doi.org/10.1016/S0021-9673(01)88619-2.

Vohland, M., M. Ludwig, S. Thiele-Bruhn, and B. Ludwig. 2014. Determination of soil proper-ties with visible to near-and mid-infrared spectroscopy: Effects of spectral variable selection. Geoderma 223:88–96. https://doi.org/10.1016/j.geoderma.2014.01.013.

Wedholm, A., L. B. Larsen, H. Lindmark-Mansson, A. H. Karlsson, and A. Andren. 2006. Effect of protein composition on the cheese-making properties of milk from individual dairy cows. J. Dairy Sci. 89:3296–3305. https://doi.org/10.3168/jds.S0022-0302(06)72366-9.

Wold, S., M. Sjöström, and L. Eriksson. 2001. PLS-Regression: A basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58:109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

Wolpert, D. H., and W. G. Macready. 1997. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1:67–82. https://doi.org/10.1109/4235.585893.

Xu, W., A. T. van Knegsel, J. J. Vervoort, R. M. Bruckmaier, R. J. van Hoeij, B. Kemp, and E. Saccenti. 2019. Prediction of metabolic status of dairy cows in early lactation with on-farm cow data and machine learning algorithms. J. Dairy Sci. 102:10186–10201. https://doi.org/10.3168/jds.2018-15791.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Series B Stat. Methodol. 67:301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

ORCIDS

M. Frizzarin ● https://orcid.org/0000-0001-7608-5504
I. C. Gormley ● https://orcid.org/0000-0001-7713-681X
D. P. Berry ● https://orcid.org/0000-0003-4349-1447
T. B. Murphy ● https://orcid.org/0000-0002-5668-7046
A. Casa ● https://orcid.org/0000-0002-2929-3850
A. Lynch ● https://orcid.org/0000-0002-8280-2933
S. McParland ● https://orcid.org/0000-0003-3320-2330