J. Dairy Sci. 105
https://doi.org/10.3168/jds.2021-21297

Predicting dry matter intake in Canadian Holstein dairy cattle using milk mid-infrared reflectance spectroscopy and other commonly available predictors via artificial neural networks

Saeed Shadpour,[1] Tatiane C. S. Chud,[1] Dagnachew Hailemariam,[2] Hinayah R. Oliveira,[1] Graham Plastow,[2] Paul Stothard,[2] Jan Lassen,[3] Ransom Baldwin,[4] Filippo Miglior,[1,5] Christine F. Baes,[1,6] Dan Tulpan,[1] and Flavio S. Schenkel[1]*

[1]Centre for Genetic Improvement of Livestock, Department of Animal Biosciences, University of Guelph, Guelph, Ontario, N1G 2W1, Canada
[2]Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, Alberta, T6G 2P5, Canada
[3]Center for Quantitative Genetics and Genomics, Aarhus University, Tjele, 8830, Denmark
[4]Animal Genomics and Improvement Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, US Department of Agriculture, Beltsville, MD 20705
[5]Lactanet Canada, Guelph, Ontario, N1K 1E5, Canada
[6]Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland 3001

ABSTRACT

Dry matter intake (DMI) is a fundamental component of the animal's feed efficiency, but measuring DMI of individual cows is expensive. Mid-infrared reflectance spectroscopy (MIRS) on milk samples could be an inexpensive alternative to predict DMI. The objectives of this study were (1) to assess if milk MIRS data could improve DMI predictions of Canadian Holstein cows using artificial neural networks (ANN); (2) to investigate the ability of different ANN architectures to predict unobserved DMI; and (3) to validate the robustness of developed prediction models. A total of 7,398 milk samples from 509 dairy cows distributed over Canada, Denmark, and the United States were analyzed. Data from Denmark and the United States were used to increase the training data size and variability to improve the generalization of the prediction models over the lactation. For each milk spectra record, the corresponding weekly average DMI (kg/d), test-day milk yield (MY, kg/d), fat yield (FY, g/d), and protein yield (PY, g/d), metabolic body weight (MBW), age at calving, year of calving, season of calving, days in milk, lactation number, country, and herd were available. The weekly average DMI was predicted with various ANN architectures using 7 predictor sets, which were created by different combinations MY, FY, PY, MBW, and MIRS data. All predictor sets also included age of calving and days in milk. In addition, the classification effects of season of calving, country, and lactation number were included in all models. The explored ANN architectures consisted of 3 training algorithms (Bayesian regularization, Levenberg-Marquardt, and scaled conjugate gradient), 2 types of activation functions (hyperbolic tangent and linear), and from 1 to 10 neurons in hidden layers). In addition, partial least squares regression was also applied to predict the DMI. Models were compared using cross-validation based on leaving out 10% of records (validation A) and leaving out 10% of cows (validation B). Superior fitting statistics of models comprising MIRS information compared with the models fitting milk, fat and protein yields suggest that other unknown milk components may help explain variation in weekly average DMI. For instance, using MY, FY, PY, and MBW as predictor variables produced a predictive accuracy (r) ranging from 0.510 to 0.652 across ANN models and validation sets. Using MIRS together with MY, FY, PY, and MBW as predictors resulted in improved fitting (r = 0.679–0.777). Including MIRS data improved the weekly average DMI prediction of Canadian Holstein cows, but it seems that MIRS predicts DMI mostly through its association with milk production traits and its utility to estimate a measure of feed efficiency that accounts for the level of production, such as residual feed intake, might be limited and needs further investigation. The better predictive ability of nonlinear ANN compared with linear ANN and partial least squares regression indicated possible nonlinear relationships between weekly average DMI and the predictor variables. In general, ANN using Bayesian regularization and scaled conjugate gradient training algorithms yielded slightly better weekly average DMI predictions compared with ANN using the Levenberg-Marquardt training algorithm.

Key words: dry matter intake, machine learning, mid-infrared reflectance spectroscopy, milk

Received September 15, 2021.
Accepted May 31, 2022.
*Corresponding author: schenkel@uoguelph.ca

INTRODUCTION

With a growing world population projected to reach 9 billion people by 2050 (UN, 2019), it is paramount to improve the efficiency of the agricultural industry. Improving feed efficiency is an important goal of many dairy producers, as feed is the main cost associated with dairy production (Hemme et al., 2014). Dry matter intake is a crucial component influencing animal efficiency, and it will be of great use for improving current genetic evaluation breeding programs (de Haas et al., 2015). However, DMI is an expensive trait to measure on individual cows in commercial systems (Dórea et al., 2018). It is recognized that the genetic evaluation of dairy cows needs fairly large data sets; however, due to the complexity and cost involved in measuring DMI, only small data sets for feed intake are currently available (Wallén et al., 2017). Consequently, the primary factor hampering the genetic evaluation of feed intake in dairy cattle breeding programs is regular access to phenotypic feed intake data from a large animal population recorded on an ongoing basis, ideally at low cost, to obtain high selection accuracy (Wallén et al., 2018). In Canada, the Efficient Dairy Genome research project (https://genomedairy.ualberta.ca/) established an international collaboration to share DMI phenotypes to enlarge the individual national databases.

Mid-infrared reflectance (**MIR**) spectroscopy (**MIRS**) on milk samples can be an affordable alternative to predict DMI at the population level (Shetty et al., 2017). Milk MIRS is based on the interaction between matter and electromagnetic waves between 900 and 5,000 cm^{-1} (De Marchi et al., 2014), and it is widely used in commercial dairy recording systems to predict milk components. In this context, MIRS has been shown to be efficient in predicting phenotypes for milk fatty acid composition (Soyeurt et al., 2006, 2011; Fleming et al., 2017), milk protein composition (Rutten et al., 2011), ketone bodies (van der Drift et al., 2012), and mineral composition (Soyeurt et al., 2009).

To the best of our knowledge, the first study on DMI prediction using milk MIRS was carried out by Shetty et al. (2017). However, in that study, the authors used the partial least squares (**PLS**) method, which may not be ideal due to the high dimensionality of the predictor variables (n = 1,060 wavelengths, **WL**), and the fact that there may be a nonlinear relationship among some WL and DMI (Dórea et al., 2018). Thus, although the PLS method has been mainly used in MIRS research to develop prediction models (e.g., Soyeurt et al., 2006; McParland et al., 2011), artificial neural networks (**ANN**) potentially may yield more accurate predictions (Gianola et al., 2011; Felipe et al., 2015). The ANN are inspired by the human nervous system, and they are comprised of interconnected neurons arranged into layers (Bishop, 2006). In this context, ANN may capture more complicated relationships between the input variables and the response outcome compared with PLS.

The objectives of this study were (1) to assess if milk MIRS data could improve prediction of DMI of Canadian Holstein cows using data from Canada and international partners; (2) to investigate the ability of different linear and nonlinear ANN architectures and PLS regression to predict DMI; and (3) to validate the robustness of developed prediction models.

MATERIALS AND METHODS

Data Collection and Quality Control

A total of 7,398 weekly average DMI from 509 first and second lactation dairy cows distributed across Canada (4,863 samples from 290 cows), the United States (1,563 samples from 127 cows), and Denmark (972 samples from 92 cows) were analyzed. Aiming at the prediction of DMI of Canadian Holstein cows only, data from Canada, and international partners (the United States and Denmark) were used to increase the data size and variability to improve the model generalization over the lactation. The data were collected between 2014 and 2018, and they were obtained from the Efficient Dairy Genome Project database (https://genomedairy.ualberta.ca/). All animal procedures were approved by the animal care and use committees at the universities and research centers that provided data for this research (i.e., the Dairy Research and Technology Center, Edmonton, Canada; the Ontario Dairy Research Centre, Elora, Canada; the Danish Cattle Research Center, Foulum, Denmark; and the USDA-Agricultural Research Service, Beltsville, MD) and all animals were cared for in accordance with respective national guidelines on animal care of each country.

All individual morning and evening milk samples were analyzed using MilkoScan FT+ spectrometers (Foss), which generated one spectrum for each analyzed milk sample. For each cow, the weekly average DMI was calculated and then it was merged with the corresponding MIRS data (see next section), based on the test date. Because some of the cow BW were originally partial BW, cow metabolic BW ($BW^{0.75}$, **MBW**) were standardized to a common mean and variance. To impute missing values of MBW, a linear regression was fit using the available MBW records for each animal and PROC GLM (version 9.4; SAS Institute Inc.). Cows without any available BW were excluded from the DMI prediction analyses. Weekly average DMI outside the range of $Q_1 - 1.5(Q_3 - Q_1)$ and $Q3 + 1.5(Q_3 - Q_1)$

Table 1. Descriptive statistics of the data, mean and SD of DMI, milk production, metabolic BW, age at calving, DIM of the recorded cows, and total number of cows and records

	Canada		United States		Denmark		All	
Variable[1]	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Weekly average DMI (kg/d)	18.70	3.76	19.36	3.58	19.84	3.49	18.99	3.71
Test-day MY (kg/d)	33.33	6.76	41.39	7.71	34.50	7.37	35.19	7.76
Test-day FY (g/d)	1,329.78	299.71	1,522.19	302.47	1,347.35	270.32	1,372.74	306.55
Test-day PY (g/d)	1,061.79	201.06	1,159.05	200.47	1,204.62	232.34	1,101.10	212.79
MBW	132.02	9.72	123.84	8.74	134.62	10.01	130.63	10.22
AC (mo)	28.01	5.99	27.38	5.71	28.92	6.01	27.99	5.95
DIM (d)	122.25	75.89	85.16	70.66	145.01	81.48	117.41	77.76
Number of records	4,863		1,563		972		7,398	
Number of cows	290		127		92		509	

[1]MY = milk yield; FY = fat yield; PY = protein yield; MBW = metabolic BW; AC = age at calving.

were considered outliers and discarded for the first and second lactation cows, where Q_1 and Q_3 represent the first and third quartiles, respectively. Descriptive statistics of the phenotypic records are shown in Table 1. The recording of DMI was focused on different stages of lactation in the United States (early lactation), Canada (post-peak of lactation) and Denmark (middle of lactation), so jointly the data sets from the 3 countries covered more of the entire lactation (Table 1).

Mid-Infrared Reflectance Spectroscopy Data and Pretreatments

For each milk sample, the milk MIRS data contained 1,060 WL in the infrared range of 925 to 5,008 cm^{-1}. Due to high water absorption, MIRS data from regions between 1,600 to 1,700 cm^{-1} and >3,005 cm^{-1} were removed (Shetty et al., 2017).

Spectral pretreatments are frequently applied to MIRS data to achieve robust prediction models (Rinnan et al., 2009). In this research, multiplicative scatter correction (Martens and Naes, 1989) was adopted to reduce noise in the spectrum. This method of preprocessing accounts for variations of light scattering in spectral data and removes nonlinearities induced by scattering from particulates (Martens and Naes, 1989). Savitzky-Golay first derivatives (Savitzky and Golay, 1964), using filter width 7 and a second order polynomial, were also applied to the spectral data. First-derivative spectra pretreatments are commonly used to enhance resolution and to sharpen bands of absorption in spectral samples (Savitzky and Golay, 1964). After the pretreatments, a total of 505 WL from each milk sample were available for the analysis.

Sets of Predictors

For each milk spectra record, the corresponding weekly average DMI (kg/d), test-day milk yield (**MY**, kg/d), fat yield (**FY**, g/d), and protein yield (**PY**, g/d), MBW, age at calving (**AC**), year of calving, season of calving, DIM, lactation number, country, and herd information were available. Seven different sets of predictors (independent variables) were designed to predict weekly average DMI of individual cows using ANN. Weekly averages were used in this study because they tend to better represent the efficiency of the cows, as DMI and other feed efficiency traits have a high daily variability (Seymour et al., 2020). Moreover, official genetic and genomic evaluations for feed efficiency in Canada are also based on weekly averages (Jamrozik and Kistemaker, 2020). The predictor sets differed by the predictors included the following: set 1 (MY, AC, and DIM), set 2 (MY, FY, PY, AC, and DIM), set 3 (MY, FY, PY, MBW, AC, and DIM), set 4 (505 WL, AC, and DIM), set 5 [MY, MBW, AC, DIM, and 36 principal components (**PC**) explaining more than 99% of MIR WL variance], set 6 (MY, FY, PY, MBW, AC, DIM, and 36 PC of MIR WL), set 7 (MY, FY, PY, MBW, AC, DIM, and 505 MIR WL). In all analyses, the fixed classification effects of country, season of calving, and lactation number were always included in the prediction models. It is worth mentioning that FY and PY were predicted values from the milk spectra. In sets 1 to 3, the simplest model (i.e., MY, AC, and DIM) was considered first, and then more predictors were added one by one but without including MIRS data. Set 4 was considered to assess how much of the DMI variation could be explained by solely using 505 MIR wavelengths. In set 5, FY and PY from set 3 were replaced by 36 PC of MIRS to evaluate whether or not MIRS data could improve DMI prediction compared with using nonspecific milk components (i.e., FY and PY). Principal components were used because in ANN, as the number of neurons increases, the number of parameters to be estimated also increases drastically, which might lead to an overfitting problem. In set 6, 36 PC of MIRS were added to the predictors of set 3

to investigate whether using PC of MIR wavelengths would or not improve DMI prediction compared with using all 505 wavelengths, as in set 7. Therefore, in set 7, all available predictors, including the full milk MIRS data were used to predict DMI. It is worth mentioning that in a study by Dórea et al. (2018), the selected MIR wavelengths (i.e., 33 MIR wavelengths), using a Markov blanket method, improved the DMI prediction by ANN compared with using all MIR wavelengths (i.e., 361 MIR wavelengths).

ANN Models

The 2-layer feed-forward perceptron, also known as single hidden layer feed-forward neural network, is a commonly applied form of ANN for regression. These ANN have an input layer of source nodes and an output unit that are fully connected, with only one hidden layer between them, as shown in Figure 1. This type of network can approximate most mathematical functions well, so they are universal approximators of linear and nonlinear functions (Alados et al., 2004). Mathematically, it is possible to view the mapping of these ANN as a 2-step regression (Hastie et al., 2009). The main idea is to extract linear combinations of the inputs as basis functions in the hidden layer, and then model the target in the output layer as a function of these basis functions (Ehret et al., 2015). In the hidden layer, the explanatory variables x_{ij} (for $j = 1, \ldots, m$, where m indicates the number of explanatory variables) of an individual i (for $i = 1, ..., n$, where n is the number of individuals) are combined linearly with a weight vector $w_{1j}^{[t]}$ determined in the training phase, plus an intercept a_t (also called "bias" in ANN terminology), with $t = 1, ..., s$ denoting the number of neurons in the hidden layer. The ensuing linear score is then converted to generate the output of the single hidden neuron $\left(z_i^{[t]}\right)$ using an activation function $f_t(.)$, as follows:

$$z_i^{[t]} = f_t\left(a_t + \sum_{j=1}^{m} w_{1j}^{[t]} x_{ij}\right), \qquad [1]$$

where all terms were previously defined.

The hyperbolic tangent activation function $\left(\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}\right)$ can be used in the hidden neurons to model the nonlinear relationship between phenotype and input, which allows the ANN to have higher flexibility than conventional linear regression models (MacKay and Mac Kay, 2003). The outputs of the basis functions, resulting from the hidden layer, are linearly combined in the output layer using the w_{21}, w_{22}, ..., w_{2s} weights and an intercept b. The resulting linear score is converted in the output neuron, this time using a linear activation function $g_t(.)$ to calculate the predicted phenotype of the individual i (for $i = 1, ..., n$), as follows:

$$y_i = g_t\left(b + \sum_{t=1}^{s} w_{2t} z_i^{[t]}\right), \qquad [2]$$

where y_i is the phenotype of the animal i, and all other terms were previously defined.

Training ANN includes minimizing an error function that relies upon the synaptic weights of the network (w in Figure 1) and biases. These weights and biases are iteratively updated to approximate the target variable by a learning algorithm. Back-propagation is a supervised learning algorithm based on an appropriate error function, the values of which are specified by the target (i.e., weekly average DMI) and the network's predicted outputs (i.e., weekly average DMI fitted values). A back-propagation algorithm is applied to determine weights and biases in a multilayer perceptron architecture to minimize the mean squared error using gradient descent methods (Okut et al., 2011). Based on the input and favored output, weights and biases in the ANN are updated continuously during the training process. In every iteration of training, an multilayer perceptron consists of 2 steps: the feed-forward and the back-propagation steps. In the feed-forward step, the inputs go through the neural network from the input layer toward the output layer and, by passing through the neurons of the hidden layer and the output layer, based on the weights, biases, and activating functions of the neural network, they yield to the neural network's outputs. The back-propagation step back propagates the errors in the network and minimizes the error function by updating weights and biases in the steepest descent direction (negative of the gradient) to decrease it (Beale et al., 2010). The forward and backward steps are carried out recurrently until the ANN solution matches the favored value within a prespecified threshold (Haykin, 2009; Hajmeer et al., 2006).

Similar to other parametric and nonparametric methods, such as kernel regression and smoothing splines, ANN may lead to overfitting and the prediction model can show low generalization (Guo et al., 2003; Feng et al., 2006; Wang et al., 2009). Overfitting results from an excessive number of estimated parameters and may happen due to an increased number of neurons in the hidden layer (Alados et al., 2004).

Bayesian regularization and cross-validated early stopping (**CVES**) are 2 standard methods to avoid overfitting in ANN. Unlike standard network training

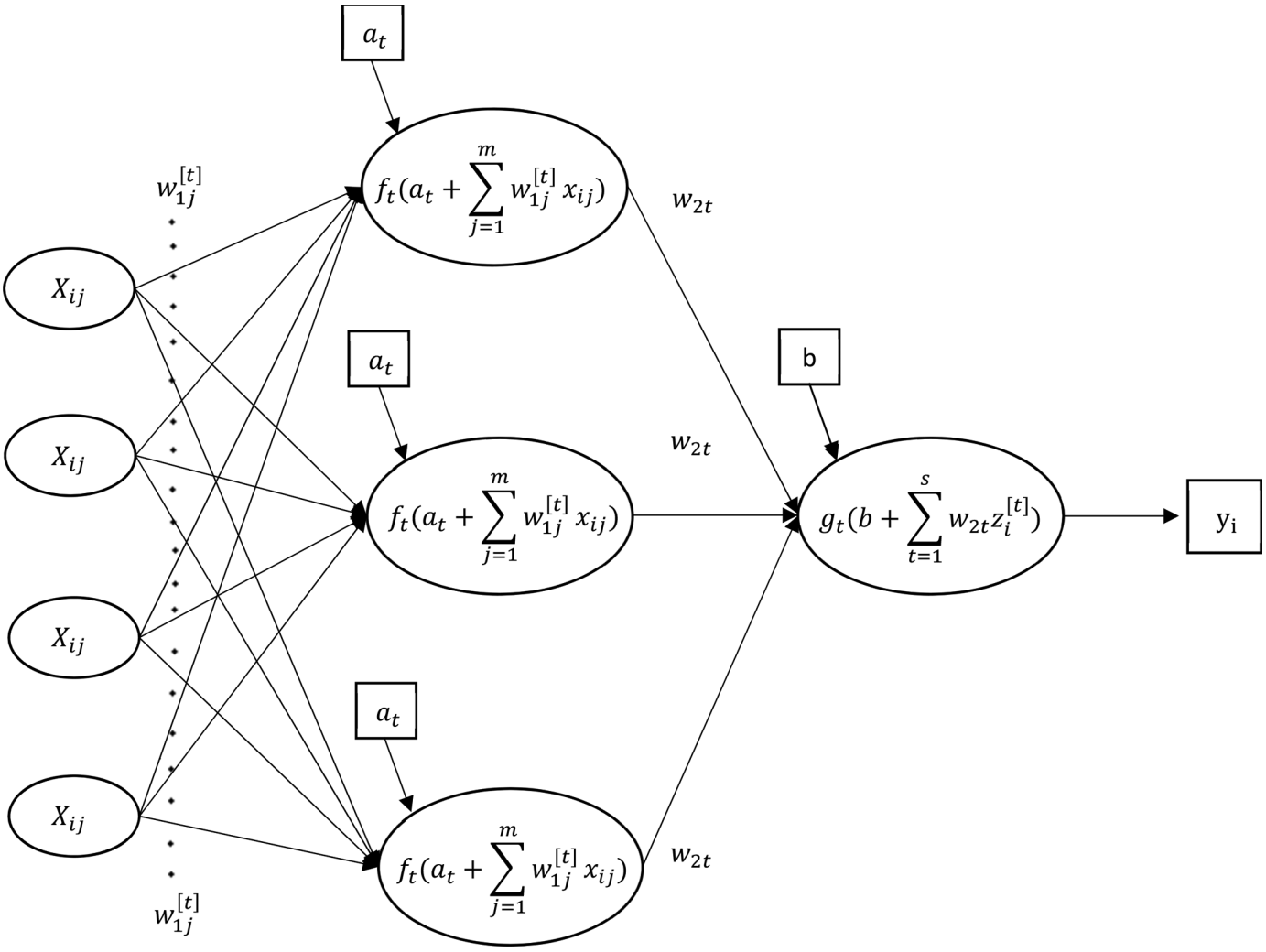

Figure 1. Architecture of a 2-layer feed-forward neural network: x_{ij} = network input; $w_{1j}^{[t]}$ = network weight from the input to hidden layer; w_{2t} = network weight from the hidden to the output layer; a_t = intercept (also called bias), with $t = 1, ..., s$, denoting the number of neurons in the hidden layer; b = the intercept of output layer neuron; y_i = network output (e.g., predicted DMI of individual); f_t (.) = activation function at the hidden neurons; g_t (.) = activation function at the output neuron, where the subscript i refers to the individuals (from 1 to n), j refers to the explanatory variables (from 1 to m), and t refers to the number of neurons (from 1 to s).

method, where an optimal set of weights and biases is selected by minimizing an error function, the Bayesian method includes a probability distribution of network weights. Consequently, the network predictions also come from a probability distribution (Sorich et al., 2003). The objective function in Bayesian ANN (e.g., BRANN) has an additional term that penalizes large weights and introduces bias in parameter estimates toward values deemed plausible, while decreasing their variance. Thus, there is a bias-variance trade-off that aims to achieve a smoother mapping (Okut et al., 2011). Afterward, gradient-based optimization is used to minimize the following function (F), which is equivalent to a penalized log-likelihood:

$$F = \beta E_D\left(D|w, M\right) + \alpha E_w\left(w|M\right), \quad [3]$$

where $E_D\left(D|w, M\right)$ and $E_w\left(w|M\right)$ are the sum of squared prediction errors and weights, respectively, considering the calibration data set (D), network weights (w), and architecture (M). The α and β are regularization parameters that must be estimated according to Foresee and Hagan, (1997). The second term on the right-hand side of Equation [3], known as weight decay, promotes low w values and alleviates the model's tendency to over-fit the data. Therefore, training involves a trade-off between model complexity and goodness of fit (Titterington, 2004).

With the CVES technique, the calibration data are divided into training and validation sets. The ANN training involves an iterative process. First, the estimation of the ANN weights and biases, using the training set, is performed and then, these estimated parameters are applied to calculate the prediction errors in the validation set. As the procedure iterates, the estimated parameters using the training data are applied to make predictions in the validation set to determine which estimated parameters yield the lowest average prediction errors for the validation data. To prevent overfitting, training stops when the error in the validation set rises in a certain number of successive epochs (iterations). The ANN parameter estimates with the best performance in the validation set are then used to assess the network′s predictive ability (Okut et al., 2013).

In this study, Levenberg-Marquardt (**LM**), scaled conjugate gradient (**SCG**) and Bayesian regularization (**BR**) training algorithms were applied to train the ANN. Additional details about LM, SCG, and BR can be found in Hagan and Menhaj (1994), Møller (1993), and MacKay (1992), respectively. The CVES method was applied for LM and SCG to avoid overfitting. Each validation set consisted of 10% of all the data chosen randomly from the calibration set. Hereafter, ANN with BR, LM, and SCG training algorithms are referred to as **BRANN**, **LMANN**, and **SCGANN** respectively.

Model Development

To fit the ANN, MATLAB (Beale et al., 2010) was used. Each network had 3 layers (i.e., input, hidden, and output layers). Two combinations of activation functions were applied: (1) a set of linear activation functions from the input layer to the hidden layer and from the hidden layer to the output layer, and (2) a set of hyperbolic tangent sigmoid activation functions from the input layer to the hidden layer, plus a linear activation function from the hidden layer to the output layer. In set 1, the number of neurons in the hidden layer was constant and equal to one, but in set 2 the number of neurons in the hidden layer varied from 1 to 10. The first and second sets of ANN were called linear and nonlinear ANN, respectively. As stated by Gianola et al. (2011) and Pérez-Rodríguez et al. (2013), linear BRANN can produce results approximately corresponding to BLUP and ridge regression. Before processing, MATLAB automatically rescales all input and output variables using the "mapminmax" function to increase the numerical stability in the range [−1, +1]. This feature scaling guarantees that all sources of data are handled equally in the training phase, which often has a large effect on the final solutions (Hastie et al., 2009).

Tuning Parameters

In machine learning, algorithms generally have a set of tuning parameters that influence how the learning algorithm fits the data. For example, the number of layers, the number of neurons in each layer, the learning rate, and other features in a multilayer perceptron ANN are tuning parameters, which are also called hyperparameters. In this study, the hyperparameters were chosen using cross-validation (i.e., goal, epochs, max_fail, min_grad, and time), and MATLAB default values (i.e., mu, mu_dec, mu_inc, mu_max, sigma, and lambda), as they are considered optimal values for most regression problems. Detailed information about the hyperparameters can be found in Table 2. For all predictor sets, the defined hyperparameters were kept constant.

Assessment of Prediction Model Robustness

Two types of 10-fold cross-validation were implemented to evaluate the predictive abilities of the ANN. The first one consisted of leaving out 10% of all animal records across Canada, Denmark, and the United States (validation A), and the second one consisted of leaving out 10% of all cows across Canada, Denmark, and the United States (validation B). The data were split into 10 subsets. One subset (the testing set) was dropped to assess the model′s predictive ability of DMI of Canadian cows only, whereas the other 9 subsets were used as calibration sets to estimate the model parameters. During the cross-validation runs, each of the 10 subsets was used once as a testing set. Such a procedure was applied to avoid having the same calibration data from the United States and Denmark in all 10-fold cross-validations, which would cause dependency in the calibration sets. Therefore, out of the total number of cows (or records) in the testing set, on average, 29 (486 records) were Canadian cows (records) and were used to evaluate the prediction ability of the models.

The predictive ability of the ANN models in the test sets was assessed based on their accuracy [r; i.e., the Pearson correlation coefficient between observed and predicted weekly average DMI values, the root mean squared error (**RMSE**), and the ratio of performance to deviation (**RPD**), which is the standard deviation of DMI divided by RMSE. Before each 10-fold cross-validation, 10 replicates using different starting values for weights and biases were applied to each ANN architecture. The starting values were chosen randomly to avoid the artificial neural network training algorithm getting stuck at the local minimum of the error function. From these 10 replicates, the estimated parameter of the replicate with the maximum Pearson correlation

coefficient between predicted and observed values in the calibration set (for BRANN) or training set (for LMANN and SCGANN) was applied to evaluate the predictive robustness of ANN in the corresponding test set.

Partial Least Squares Regression

For the sake of comparison to previous studies, PLS regression was also applied to predict the DMI, using the PLS package in R (Mevik et al., 2013). The optimum number of latent variables for the PLS regression model for each 10-fold cross-validation was determined using the maximum Pearson correlation between observed and predicted DMI in the validation sets (i.e., the 10% of all the data chosen randomly from the calibration set) when the number of latent variables was changed from 1 to 100 for sets 4 and 7 (to reduce the number of latent variables), and from 1 to the number of predictors for the other sets. Then the determined optimum number of latent variables was used to make a prediction model using the corresponding calibration data (i.e., both training and validation data) for each 10-fold cross-validation.

Statistical Tests

Differences in fitting statistics (i.e., r, RMSE and RPD) among 7 sets of predictors (i.e., predictor sets), 3 types of training algorithms (i.e., BR, LM, and SCG), linear and nonlinear ANN and PLS regression, and 2 types of validation methods (i.e., validations A and B) were tested via SAS PROC GLM, using Scheffé adjustment for multiple comparisons. The ANOVA model included 10-fold samples as a random factor (i.e., blocking for 10-fold samples) and predictor sets, models (i.e., linear and nonlinear ANN and PLS regression), training algorithms, validation methods, and predictor sets as fixed categorical factors.

RESULTS

The comparisons between predictor sets, models, and training algorithms were very similar within validation method A or B, but validation B yielded consistently lower fitting statistics than validation A ($P < 0.0001$; Supplemental Table S1, https://doi.org/10.5281/zenodo.6632059; Shadpour et al., 2022). As validation B is more practical, because DMI of new animals are predicted, only results from validation B are presented and discussed hereafter, with detailed statistical analyses provide in Supplemental Table S2 (https://doi.org/10.5281/zenodo.6632059; Shadpour et al., 2022). Results from validation A are provided in the Supplemental Tables S3 and S4 (https://doi.org/10.5281/zenodo.6632059; Shadpour et al., 2022). In addition, only statistical comparisons for prediction accuracy are presented, because results for RMSE and RPD followed similar trends.

Data Description

The mean phenotypic values of the different predictor sets for the 509 cows, on days where MIRS data were available, are summarized in Table 1. Weekly average DMI and its corresponding test-day MY, FY,

Table 2. Hyperparameters used in the artificial neural networks (ANN) based on Levenberg-Marquardt (LMANN), scaled conjugate gradient (SCGANN), and Bayesian regularization (BRANN) training algorithms

	Linear ANN			Nonlinear ANN		
Hyperparameter[1]	BRANN	LMANN	SCGANN	BRANN	LMANN	SCGANN
goal	0	0	0	0	0	0
epochs	70	70	10,000	50	50	1,000
max_fail	—	6	3,000	—	6	300
min_grad	1e-100	1e-100	1e-100	1e-100	1e-100	1e-100
time	infinite	infinite	infinite	infinite	infinite	infinite
mu	0.005	0.001	—	0.005	0.001	—
mu_dec	0.1	0.1	—	0.1	0.1	—
mu_inc	10	10	—	10	10	—
mu_max	1e10	1e10	—	1e10	1e10	—
sigma	—	—	5.0e-5	—	—	5.0e-5
lambda	—	—	5.0e-7	—	—	5.0e-7

[1]goal = performance goal; epochs = maximum number of epochs to train; max_fail = maximum validation failures; min_grad = minimum performance gradient; time = maximum time to train in seconds; mu = initial momentum; mu_dec = mu decrease factor; mu_inc = mu increase factor; mu_max = maximum mu; sigma = determine change in weight for second derivative approximation; lambda = parameter for regulating the indefiniteness of the Hessian. The first 5 hyperparameters (i.e., goal, epochs, max_fail, min_grad, and time) were chosen using cross-validation. The other hyperparameters (i.e., mu, mu_dec, mu_inc, mu_max, sigma, and lambda) were MATLAB default values, which are optimal for most regression problems.

Table 3. Fitting statistics of the artificial neural networks (ANN) using Bayesian regularization training algorithm (BRANN) for predicting weekly average DMI in Canadian Holstein dairy cattle using the validation B (based on individual cows) in the test sets[1]

Set[2]	Predictive ability	L	NN1	NN2	NN3	NN4	NN5	NN6	NN7	NN8	NN9	NN10
1	r	0.471	0.526	0.529	0.546	0.533	0.537	0.537	0.530	0.539	0.534	0.532
	RMSE	3.332	3.230	3.212	3.165	3.203	3.195	3.205	3.217	3.201	3.217	3.227
	RPD	1.114	1.149	1.156	1.172	1.159	1.162	1.158	1.154	1.160	1.153	1.150
2	r	0.488	0.531	0.556	0.554	0.554	0.563	0.572	0.581	0.576	0.575	0.569
	RMSE	3.314	3.220	3.145	3.142	3.144	3.123	3.101	3.087	3.091	3.105	3.120
	RPD	1.120	1.154	1.181	1.181	1.181	1.188	1.197	1.202	1.200	1.194	1.189
3	r	0.510	0.543	0.576	0.577	0.572	0.575	0.593	0.577	0.591	0.588	0.577
	RMSE	3.267	3.194	3.104	3.095	3.104	3.099	3.047	3.097	3.065	3.077	3.121
	RPD	1.136	1.162	1.196	1.198	1.196	1.199	1.218	1.197	1.210	1.206	1.190
4	r	0.636	0.637	0.641	0.653	0.65	0.654	0.654	0.651	0.634	0.656	0.635
	RMSE	2.915	2.909	2.905	2.863	2.893	2.895	2.875	2.924	2.964	2.909	2.983
	RPD	1.273	1.275	1.276	1.296	1.282	1.283	1.292	1.268	1.253	1.275	1.244
5	r	0.652	0.655	0.663	0.682	0.691	0.714	0.710	0.706	0.711	0.714	0.698
	RMSE	2.866	2.860	2.833	2.770	2.732	2.643	2.674	2.696	2.665	2.649	2.724
	RPD	1.293	1.297	1.309	1.340	1.357	1.404	1.388	1.377	1.391	1.402	1.364
6	r	0.658	0.662	0.672	0.693	0.692	0.725	0.716	0.713	0.705	0.707	0.705
	RMSE	2.845	2.832	2.805	2.719	2.723	2.600	2.640	2.655	2.685	2.683	2.708
	RPD	1.304	1.310	1.322	1.365	1.363	1.427	1.407	1.400	1.384	1.386	1.372
7	r	0.689	0.691	0.692	0.705	0.719	0.729	0.721	0.720	0.708	0.728	0.725
	RMSE	2.740	2.728	2.734	2.699	2.650	2.601	2.625	2.630	2.688	2.616	2.625
	RPD	1.354	1.361	1.357	1.374	1.400	1.426	1.414	1.411	1.383	1.421	1.415

[1]L = linear neural network; NN = number of neurons in the hidden layer of nonlinear neural networks. The results shown in this table are averages based on a 10-fold cross-validation.

[2]Sets of explanatory variables used are defined as follows: (1) MY, AC, and DIM; (2) MY, FY, PY, AC, and DIM; (3) MY, FY, PY, MBW, AC, and DIM; (4) AC, DIM, and 505 WL; (5) MY, MBW, AC, DIM, and 36 principal components of WL; (6) MY, FY, PY, MBW, AC, DIM, and 36 principal components of WL; and (7) MY, FY, PY, MBW, AC, DIM, and 505 WL, where MY = milk yield, AC = age at calving, FY = fat yield, PY = protein yield, MBW = metabolic BW, and WL = wavelength. All sets also included country, season of calving, and lactation number as categorical fixed effects. The predictive ability of the BRANN was assessed using the Pearson correlation coefficient calculated between observed and predicted DMI values (r), the root mean squared error (RMSE) and the ratio of performance to deviation (RPD).

PY, MBW, AC, and DIM were 18.99 kg, 35.19 kg, 1,372.7 g, 1,101.1 g, 130.63 $kg^{0.75}$, 28.0 mo, and 117.4 d, respectively.

Common Categorical Environmental Effects Across Sets

In all the analyzed sets (sets 1–7), country, season of calving, and lactation number were always included as categorical fixed environmental effects. Their contributions to the prediction accuracy were small. For instance, when using BRANN and Validation A, their joint contribution varied from 0.014 to 0.075. The same features for validation B were from −0.012 to 0.014.

Comparison Between Predictor Sets, Models, and Training Algorithms Using Validation B

The fitting statistics of BRANN, LMANN, and CSGANN for Canadian Holstein dairy cattle using the validation B are presented in Tables 3 to 5, respectively. Predicting weekly average DMI using MY as the only production trait (set 1) resulted in the worst fitting statistics compared with the other predictor sets ($P < 0.0001$, Supplemental Table S2). The predictive accuracy for set 1 ranged from 0.471 to 0.546 for BRANN, 0.472 to 0.55 for LMANN, and 0.473 to 0.546 for SCGANN. The addition of FY and PY (set 2) improved the prediction performance ($P < 0.0001$). The predictive accuracy ranged from 0.488 to 0.581 for BRANN, 0.489 to 0.567 for LMANN, and 0.484 to 0.58 for SCGANN.

When MBW was included in the prediction model (set 3), the fitting statistics improved ($P < 0.0001$). The predictive accuracy ranged from 0.51 to 0.593 for BRANN, 0.512 to 0.591 for LMANN, and 0.511 to 0.591 for SCGANN.

When milk production traits were replaced by milk spectra (505 WL) in set 4, predictive accuracy considerably increased ($P < 0.0001$) and ranged from 0.634 to 0.656 for BRANN, 0.603 to 0.645 for LMANN, and 0.635 to 0.652 for SCGANN.

Replacing FY and PY by the PC of WL (set 5) resulted in a larger improvement in the fitting statistics ($P < 0.0001$). The predictive accuracy ranged from 0.652 to 0.714 for BRANN, 0.653 to 0.703 for LMANN, and 0.652 to 0.709 for SCGANN. Conversely, combining FY and PY with MY, SMBY and the PC of WL (set 6) did not have a major effect on the DMI prediction compared with set 5 ($P = 0.57$). The predictive accuracy ranged from 0.658 to 0.725 for BRANN, 0.659 to 0.713 for LMANN, and 0.658 to 0.707 for SCGANN.

Table 4. Fitting statistics of the artificial neural networks (ANN) using Levenberg-Marquardt training algorithm (LMANN) for predicting weekly average DMI in Canadian Holstein dairy cattle using the validation B (based on individual cows) in the test sets[1]

Set[2]	Predictive ability	L	NN1	NN2	NN3	NN4	NN5	NN6	NN7	NN8	NN9	NN10
1	r	0.472	0.525	0.531	0.542	0.550	0.529	0.529	0.532	0.527	0.540	0.517
	RMSE	3.333	3.230	3.208	3.176	3.157	3.209	3.217	3.212	3.225	3.203	3.265
	RPD	1.114	1.149	1.157	1.168	1.175	1.157	1.154	1.155	1.151	1.160	1.138
2	r	0.489	0.532	0.556	0.558	0.552	0.558	0.567	0.555	0.553	0.553	0.560
	RMSE	3.305	3.217	3.141	3.131	3.157	3.125	3.107	3.144	3.159	3.162	3.137
	RPD	1.123	1.155	1.183	1.185	1.176	1.189	1.195	1.181	1.174	1.172	1.184
3	r	0.512	0.542	0.575	0.577	0.567	0.578	0.591	0.581	0.574	0.565	0.558
	RMSE	3.262	3.195	3.103	3.097	3.115	3.098	3.058	3.116	3.115	3.157	3.159
	RPD	1.137	1.162	1.196	1.198	1.192	1.197	1.214	1.191	1.191	1.175	1.175
4	r	0.643	0.645	0.633	0.635	0.631	0.625	0.617	0.625	0.634	0.603	0.619
	RMSE	2.897	2.889	2.934	2.95	2.978	2.989	3.044	3.038	3.02	3.14	3.097
	RPD	1.281	1.284	1.265	1.258	1.246	1.243	1.223	1.221	1.227	1.183	1.199
5	r	0.653	0.656	0.665	0.672	0.697	0.700	0.703	0.695	0.696	0.694	0.694
	RMSE	2.864	2.854	2.827	2.798	2.714	2.714	2.694	2.729	2.709	2.724	2.765
	RPD	1.295	1.300	1.313	1.326	1.367	1.366	1.377	1.361	1.370	1.362	1.341
6	r	0.659	0.663	0.675	0.686	0.698	0.713	0.704	0.702	0.699	0.703	0.703
	RMSE	2.843	2.831	2.793	2.748	2.706	2.661	2.697	2.695	2.724	2.696	2.716
	RPD	1.305	1.311	1.328	1.350	1.372	1.395	1.377	1.379	1.361	1.377	1.368
7	r	0.695	0.703	0.690	0.702	0.703	0.699	0.679	0.680	0.679	0.691	0.682
	RMSE	2.719	2.685	2.754	2.709	2.706	2.753	2.827	2.848	2.824	2.806	2.832
	RPD	1.365	1.382	1.349	1.370	1.369	1.350	1.312	1.308	1.315	1.324	1.313

[1]L = linear neural network; NN = number of neurons in the hidden layer of nonlinear neural networks. The results shown in this table are averages based on a 10-fold cross-validation.

[2]Sets of explanatory variables used are defined as follows: (1) MY, AC, and DIM; (2) MY, FY, PY, AC, and DIM; (3) MY, FY, PY, MBW, AC, and DIM; (4) AC, DIM, and 505 WL; (5) MY, MBW, AC, DIM, and 36 principal components of WL; (6) MY, FY, PY, MBW, AC, DIM, and 36 principal components of WL; and (7) MY, FY, PY, MBW, AC, DIM, and 505 WL; where MY = milk yield, AC = age at calving, FY = fat yield, PY = protein yield, MBW = metabolic BW, and WL = wavelength. All sets also included country, season of calving, and lactation number as categorical fixed effects. The predictive ability of the LMANN was assessed using the Pearson correlation coefficient calculated between observed and predicted DMI values (r), the root mean squared error (RMSE), and the ratio of performance to deviation (RPD).

Table 5. Fitting statistics of the artificial neural networks (ANN) using conjugate gradient training algorithm (SCGANN) for predicting weekly average DMI in Canadian Holstein dairy cattle using the validation B (based on individual cows) in the test sets[1]

Set[2]	Predictive ability	L	NN1	NN2	NN3	NN4	NN5	NN6	NN7	NN8	NN9	NN10
1	r	0.473	0.525	0.527	0.546	0.545	0.532	0.536	0.539	0.530	0.526	0.539
	RMSE	3.331	3.230	3.216	3.165	3.176	3.200	3.201	3.193	3.219	3.246	3.197
	RPD	1.114	1.149	1.155	1.172	1.169	1.159	1.160	1.162	1.154	1.143	1.160
2	r	0.484	0.532	0.557	0.554	0.563	0.565	0.580	0.566	0.580	0.561	0.562
	RMSE	3.315	3.217	3.141	3.140	3.125	3.117	3.090	3.126	3.088	3.142	3.145
	RPD	1.120	1.155	1.183	1.182	1.188	1.191	1.201	1.188	1.202	1.181	1.181
3	r	0.511	0.543	0.575	0.577	0.577	0.584	0.591	0.578	0.574	0.584	0.584
	RMSE	3.261	3.193	3.102	3.091	3.083	3.072	3.050	3.097	3.097	3.092	3.084
	RPD	1.137	1.163	1.197	1.200	1.204	1.208	1.218	1.198	1.198	1.200	1.204
4	r	0.640	0.635	0.642	0.645	0.652	0.652	0.641	0.65	0.646	0.643	0.635
	RMSE	2.906	2.917	2.903	2.914	2.891	2.898	2.945	2.922	2.94	2.98	3.005
	RPD	1.277	1.272	1.277	1.272	1.284	1.279	1.258	1.269	1.261	1.246	1.235
5	r	0.652	0.655	0.661	0.686	0.695	0.701	0.694	0.709	0.693	0.690	0.700
	RMSE	2.865	2.856	2.844	2.751	2.715	2.692	2.735	2.682	2.731	2.757	2.703
	RPD	1.294	1.298	1.305	1.348	1.367	1.381	1.357	1.384	1.360	1.346	1.374
6	r	0.658	0.661	0.675	0.689	0.697	0.701	0.702	0.706	0.692	0.707	0.703
	RMSE	2.847	2.834	2.782	2.743	2.723	2.712	2.705	2.682	2.743	2.693	2.695
	RPD	1.303	1.309	1.332	1.352	1.363	1.369	1.374	1.383	1.352	1.379	1.377
7	r	0.692	0.692	0.695	0.700	0.711	0.717	0.717	0.705	0.726	0.720	0.710
	RMSE	2.730	2.727	2.726	2.703	2.673	2.655	2.650	2.697	2.626	2.655	2.692
	RPD	1.359	1.361	1.361	1.374	1.387	1.395	1.399	1.375	1.415	1.395	1.380

[1]L = linear neural network; NN = number of neurons in the hidden layer of nonlinear neural networks. The results shown in this table are averages based on a 10-fold cross-validation.

[2]Sets of explanatory variables used are defined as follows: (1) MY, AC, and DIM; (2) MY, FY, PY, AC, and DIM; (3) MY, FY, PY, MBW, AC, and DIM; (4) AC, DIM, and 505 WL; (5) MY, MBW, AC, DIM, and 36 principal components of WL; (6) MY, FY, PY, MBW, AC, DIM, and 36 principal components of WL; and (7) MY, FY, PY, MBW, AC, DIM, and 505 WL; where MY = milk yield, AC = age at calving, FY = fat yield, PY = protein yield, MBW = metabolic BW, and WL = wavelength. All sets also included country, season of calving, and lactation number as categorical fixed effects. The predictive ability of the SCGANN was assessed using the Pearson correlation coefficient calculated between observed and predicted DMI values (r), the root mean squared error (RMSE), and the ratio of performance to deviation (RPD).

Using all WL combined with MY, FY, PY, and MBW (i.e., set 7) resulted in the best fitting statistics compared with the other sets ($P < 0.0001$). The predictive accuracy ranged from 0.689 to 0.729 for BRANN, 0.679 to 0.703 for LMANN, and 0.692 to 0.726 for SCGANN.

For all predictor sets, nonlinear ANN provided better prediction ability compared with linear ANN. PLS regression yielded very similar results compared with linear ANN ($P = 0.99$, Table 6 and Supplementary Table S2). For all sets, the predictive abilities of the 3 neural network training algorithms were similar, but the BRANN and SCGANN performed slightly better than LMANN ($P < 0.0001$).

In this study, data from Canada, Denmark and the United States were combined to increase the training data size and variability, and improve the generalization of the prediction models over the lactation. As recording of DMI was focused on different stages of lactation in the United States, Canada, and Denmark (average DIM of 85 d, 122 d, and 145 d, respectively, with standard deviations ranging from 71 to 81 d; Table 1), jointly the data sets from the 3 countries covered more of the entire lactation.

The prediction ability of Canadian cows' DMI using data from Canada, Denmark, and the United States or only data from Canada was very similar, but it was, on average, slightly better when all 3 countries were used. As an illustration, Table 7 shows the average fitting statistics over all the 11 tested ANN using BRANN for predicting the weekly average DMI of Canadian cows using predictor set 7, validation B and data of either Canada only or Canada, Denmark, and the United States combined in the test sets for the whole lactation or different lactation periods. The results show a slightly better overall fitting of the models when using the combined data sets and a slightly improved fitting in early and later lactation when compared with using Canadian data only. These results supported the use of combined data from Canada, the United States, and Denmark to develop DMI predictions of Canadian cows.

Table 6. Fitting statistics of the partial least squares (PLS) regression model for predicting weekly average DMI in Canadian Holstein dairy cattle using the validation B (based on individual cows) in the test sets

Set[1]	Predictive ability[2]	PLS
1	r	0.467
	RMSE	3.342
	RPD	1.111
2	r	0.486
	RMSE	3.313
	RPD	1.120
3	r	0.506
	RMSE	3.274
	RPD	1.133
4	r	0.637
	RMSE	2.916
	RPD	1.273
5	r	0.652
	RMSE	2.867
	RPD	1.293
6	r	0.658
	RMSE	2.845
	RPD	1.304
7	r	0.665
	RMSE	2.821
	RPD	1.315

[1]Sets of explanatory variables used are defined as follows: (1) MY, AC, and DIM; (2) MY, FY, PY, AC, and DIM; (3) MY, FY, PY, MBW, AC, and DIM; (4) AC, DIM, and 505 WL; (5) MY, MBW, AC, DIM, and 36 principal components of WL; (6) MY, FY, PY, MBW, AC, DIM, and 36 principal components of WL; and (7) MY, FY, PY, MBW, AC, DIM, and 505 WL; where MY = milk yield, AC = age at calving, FY = fat yield, PY = protein yield, MBW = metabolic BW, and WL = wavelength. All sets also included country, season of calving, and lactation number as categorical fixed effects.

[2]The predictive ability of the PLS regression was assessed using the Pearson correlation coefficient calculated between observed and predicted DMI values (r), the root mean squared error (RMSE), and the ratio of performance to deviation (RPD). The results shown in this table are averages based on a 10-fold cross-validation.

DISCUSSION

The objective of the present study was to assess the prediction of weekly average DMI in Canadian Holstein dairy cattle using milk MIRS data in addition to other commonly available predictors, including MY, milk components and MBW, using ANN and PLS, and combined data from Canada and 2 international partners, Denmark and the United States. Accurate prediction of

Table 7. Average fitting statistics over all 11 tested artificial neural networks (ANN) using the Bayesian regularization (BRANN) training algorithm for predicting weekly average DMI in Canadian Holstein dairy cattle in different lactation periods using predictor set 7, validation B (based on individual cows)[1]

		Test set	
Statistic[2]	DIM	CA	All
r	5–305	0.708	0.712
	5–90	0.702	0.712
	91–180	0.604	0.593
	181–305	0.503	0.506
RMSE	5–305	2.702	2.667
	5–90	2.727	2.665
	91–180	2.607	2.626
	181–305	2.758	2.696
RDP	5–305	1.376	1.392
	5–90	1.356	1.389
	91–180	1.228	1.217
	181–305	1.089	1.110

[1]Test sets included data from either Canada only (CA) or Canada, Denmark, and the United States together (All).

[2]The predictive ability of the BRANN was assessed using the Pearson correlation coefficient calculated between observed and predicted DMI values (r), the root mean squared error (RMSE), and the ratio of performance to deviation (RPD).

weekly average DMI of individual cows can assist feed cost management and allow the genetic improvement for feed performance by providing large-scale data for prediction of breeding values. As stated by Wallén et al. (2018), at least 4,000 genotyped and phenotyped heifers need to be added annually to the reference population to genetically improve feed efficiency using genomic selection. Providing such large number of animals with feed intake measures at low cost is challenging. One possible solution to overcome this conundrum is the indirect measurement of feed intake from other available data, such as MIRS. Therefore, it is of particular interest to evaluate the predictive ability of the different combinations of currently available predictor sets, as well as the usefulness of ANN models and PLS.

Cross-Validation to Assess Model Robustness

Cross-validation was used to identify the optimal predictor set and ANN architecture, and to evaluate the robustness of the prediction models. The prediction models for weekly average DMI were evaluated using 2 types of 10-fold cross-validation: validation on individual records (A) and animals (B). In validation A, records of the same cow were presented in both the calibration and test sets, which likely led to an inflation in the model predictions. In this context, a more realistic situation was considered in validation B, where 10% of the cows were removed from the calibration data in each of the 10-folds. Therefore, validation B assessed how the model would perform when predicting weekly average DMI for new cows. Data dependencies between calibration and test sets, which were reported when animals from the same herds are included in both calibration and test data sets, may inflate prediction performance (e.g., Wang and Bovenhuis, 2019; Coelho Ribeiro et al., 2021). However, this is not always the case, as in study by Lahart et al. (2019), where the average accuracy of DMI prediction in test sets using only MIRS and MIRS combined with MY, F%, P%, BW, stage of lactation and parity, using within herd and across-herd cross-validations, were 0.69, 0.87, 0.55, and 0.80, respectively. There are some aspects in the current study that might suggest that the results were not noticeably affected by data dependencies between calibration and test sets. None of the 7 sets of predictive variables included the herd effect. Predicting DMI using only the herd effect resulted in very poor prediction performance (results not shown). This suggests that the herd effect does not have a major contribution to the DMI prediction. In addition, the adapted BRANN used regularization to handle overfitting. Regularization penalizes large weights and delivers a smoother fit. Consequently, in the regularized models, all features contribute to making a prediction, which almost guarantees that the final prediction does not come from dependencies between calibration and test sets (i.e., all features cannot have major dependencies between calibration and test sets at the same time).

Predictor Sets

The predictive performance of 7 different combinations of predictor sets was investigated in this study. Wallén et al. (2018) evaluated DMI prediction in lactating Norwegian Red dairy cows using PLS and MIRS data and reported lower prediction accuracy compared with the present study when using MIRS data to predict weekly average DMI. Adding milk components (i.e., FY and PY) to MY enhanced the weekly average DMI predictions, which agreed with the results of Shetty et al. (2017) and Wallén et al. (2017). Our results support the hypothesis that MBW contributes to explaining DMI variation in dairy cattle, because the prediction ability improved when MBW was added to the model. These findings are in line with the results of previous studies in dairy cattle (e.g., Shetty et al., 2017; Dórea et al., 2018; Wallén et al., 2018). The improvement in fitting statistics when MIRS information was combined with the other predictors (i.e., sets 4–7) indicates that MIRS provides some additional information to predict weekly average DMI. In this context, it should be highlighted that including all MIRS WL (i.e., set 7) resulted in better prediction of weekly average DMI compared with the predictors sets that included the PC of WL (i.e., sets 5 and 6). However, when only MIRS WL were included in the model (set 4) the best prediction accuracy was slightly higher than when all production traits were included in the model (set 2), showing that some of the information from MIRS may be captured through its association with production traits.

As previously mentioned, the inclusion of FY, PY, and MBW in the predictor set including MY led to a substantial improvement in the weekly average DMI prediction (set 1 vs. set 3). In addition, there was a further improvement in the fitting statistics when PC of WL were included in the model (set 6). However, removing FY and PY when PC of WL were included (set 5) did not have a major effect on the fitting statistics. These findings might be related to the fact that WL can fairly describe milk fatty acid and protein profiles, as it contains more detailed information on fatty acids (Soyeurt et al., 2011) and proteins (McDermott et al., 2016) compared with nonspecific fat and protein contents used in set 3. Similar findings on DMI prediction using MIRS data have been reported by Lahart et al. (2019) for grazing dairy cows. Significant improvement in the DMI prediction of TMR-fed cows have also been

reported by Dórea et al. (2018) when ANN and milk MIRS data were used. Nonetheless, Shetty et al. (2017) did not report any superior DMI predictive ability for models that included MIRS instead of FY and PY, on cows under a TMR diet. Additionally, Wallén et al. (2018) reported that including only MY and BW in the prediction model resulted in similar accuracy as including MIRS data.

The prediction accuracies of weekly average DMI using MIRS data in this study were lower than those of Shetty et al. (2017), Dórea et al. (2018), and Lahart et al. (2019), but higher than the ones reported by Wallén et al. (2018). Differences in accuracies across studies may be due to the differences in sample sizes used to develop the prediction models, frequency of recording, breeds, diets, and prediction methods, as well as using different WL, training, and validation data sets. Using daily or weekly average milk MIRS data as predictors can also affect the prediction accuracy (Shetty et al.,2017).

Comparison Between Linear and Nonlinear ANN and PLS

The ANN are powerful and flexible tools to model potentially nonlinear relationships between input and output variables. In this study and others (e.g., Perai et al., 2010; Lin et al., 2012; Dórea et al., 2018), ANN provided better predictions than linear models. Nonlinear ANN yielded superior prediction than linear ANN and PLS. Moreover, the superiority of nonlinear ANN was more evident in predictor sets that did not include MIRS information (i.e., sets 1–3). This agrees with Dórea et al. (2018), who applied ANN and PLS for the DMI prediction using MIRS data. These results suggest that some nonlinear interactions exist between the weekly average DMI and the predictors used.

Comparison Between BRANN, LMANN, and SCGANN

In the present study, 3 types of training algorithms, including BR, LM, and SCG, were applied to train the ANN. Furthermore, to attenuate overfitting and obtain models that generalize well, 2 strategies were used: regularization (for BRANN), and CVES (for both LMANN and SCGANN). In general, BRANN and SCGANN yielded better weekly average DMI predictions compared with LMANN. Bayesian regularization showed nonsignificant, slightly better predictions than SCGANN, but significantly better predictions in validation A (data not shown) compared with validation B. Okut et al. (2013) also reported that BRANN over performed SCGANN when predicting the expected progeny difference for marbling score in Angus cattle using ANN and 2 sets of SNP panels.

Usefulness in Practical Applications

Predictors for weekly average DMI prediction, which include MY, milk components (i.e., FY and PY), and to a lesser extent MBW, are available for many cows. Therefore, the question of interest is whether MIRS information can add additional information that is not currently available. The results from the current study showed that the best predictor set for weekly average DMI prediction consisted of MY, FY, PY, MBW, and full MIRS data (predictor set 7). Thus, MIRS information increased the prediction accuracy for weekly average DMI. For linear ANN, the prediction accuracy of predictor set 7 was on average 0.19 points higher than set 3 over all prediction methods and validation data sets, whereas the same feature for the best nonlinear ANN was 0.12. This means that if MIRS data are stored, it is possible to achieve more accurate prediction of weekly average DMI. However, it is worth recalling that the best prediction accuracies were obtained using nonlinear ANN, regardless of the fact that MIRS contributed less to the overall accuracy of these models.

A further consideration is the practical application of the predictor set, because weekly average DMI prediction methods must be easily applicable to commercial dairy farms. Most variables in the predictor sets used in this study are readily available, with the possible exception of BW. Although BW is not accessible on many commercial dairy farms, it can be estimated using, for example, chest width (Veerkamp and Brotherstone, 1997). Milk MIRS data are ideally suited to predict DMI without additional cost, because it has been currently used to measure milk components of milk-recorded animals (McParland et al., 2014; Shetty et al., 2017).

As noted before, when only MIR WL were included in the model (set 4) the best prediction accuracy was slightly higher than that when all production traits were included in the model (set 2). It should be mention that a part of the DMI variation explained by MIR may be captured through its association with production traits. This might limit the application of MIR to estimating feed efficiency measurements that accounts for the level of production, such as residual feed intake. Development in sensors technology could provide the opportunity to include more informative predictors in the prediction models, such as weather data, cow's body temperature, feeding behavior (Dolecheck et al., 2016; Borchers et al., 2017), herbage mass and allowance in grazing dairy cows (van Knegsel et al., 2010; Muñoz et

al., 2016). Therefore, further improvements in weekly average DMI prediction may be achieved when accurate data on such predictors become available.

CONCLUSIONS

The model including known animal energy sinks (production traits and MBW), MIRS, and environmental predictors showed the best prediction of weekly average DMI in Canadian Holstein cows, especially when nonlinear ANN were applied. The superior fitting statistics of models including MIRS compared with models without MIRS suggest that other unknown milk components may help explain variation in weekly average DMI. Nevertheless, only a marginal gain in accuracy was observed when production traits and MBW were included in the model after MIRS. However, it seems that MIRS predicts DMI mostly through its association with milk production traits and its utility in estimating a measure of feed efficiency that accounts for the level of production, such as residual feed intake, might be limited and would need further investigation.

ACKNOWLEDGMENTS

The authors acknowledge the financial support from the Efficient Dairy Genome Project funded by Genome Canada (Ottawa, ON, Canada), Genome Alberta (Calgary, AB, Canada), Ontario Genomics (Toronto, ON, Canada), Alberta Agriculture and Forestry (Edmonton, AB, Canada), Ontario Ministry of Research and Innovation (Toronto, ON, Canada), Ontario Ministry of Agriculture, Food and Rural Affairs (Guelph, ON, Canada), Canadian Dairy Network (Guelph, ON, Canada), GrowSafe Systems (Airdrie, AB, Canada), Alberta Milk (Edmonton, AB, Canada), Victoria Agriculture (Melbourne, Australia), Scotland's Rural College (Edinburgh, UK), USDA Agricultural Research Service (Beltsville, MD), Qualitas AG (Zug, Switzerland), and Aarhus University (Aarhus, Denmark). C. F. Baes and F. S. Schenkel also acknowledge financial contributions from the Dairy Research Cluster 3 (Dairy Farmers of Canada and Agriculture and Agri-Food Canada) under the Canadian Agricultural Partnership AgriScience Program, and from the Efficient Dairy Genome Project. The authors have not stated any conflicts of interest.

REFERENCES

Alados, I., J. A. Mellado, F. Ramos, and L. Alados-Arboledas. 2004. Estimating UV erythemal irradiance by means of neural networks. Photochem. Photobiol. 80:351–358. https://doi.org/10.1562/2004-03-12-RA-111.1.

Beale, M. H., M. T. Hagan, and H. B. Demuth. 2010. Neural network toolbox. User's Guide. MathWorks 2:77–81.

Bishop, C. M. 2006. Neural networks. Pages 227–281 in Pattern Recognition and Machine Learning. Springer.

Borchers, M. R., Y. M. Chang, K. L. Proudfoot, B. A. Wadsworth, A. E. Stone, and J. M. Bewley. 2017. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. J. Dairy Sci. 100:5664–5674. https://doi.org/10.3168/jds.2016-11526.

Coelho Ribeiro, L. A., T. Bresolin, G. J. de Magãlhaes Rosa, D. Rume Casagrande, M. A. C. Danes, and J. R. R. Dórea. 2021. Disentangling data dependency using cross-validation strategies to evaluate prediction quality of cattle grazing activities using machine learning algorithms and wearable sensor data. J. Anim. Sci. 99:skab206.

de Haas, Y., J. E. Pryce, M. P. L. Calus, E. Wall, D. P. Berry, P. Løvendahl, N. Krattenmacher, F. Miglior, K. Weigel, D. Spurlock, K. A. Macdonald, B. Hulsegge, and R. F. Veerkamp. 2015. Genomic prediction of dry matter intake in dairy cattle from an international data set consisting of research herds in Europe, North America, and Australasia. J. Dairy Sci. 98:6522–6534. https://doi.org/10.3168/jds.2014-9257.

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.

Dolecheck, K. A., W. J. Silvia, G. Heersche Jr., C. L. Wood, K. J. McQuerry, and J. M. Bewley. 2016. A comparison of timed artificial insemination and automated activity monitoring with hormone intervention in 3 commercial dairy herds. J. Dairy Sci. 99:1506–1514. https://doi.org/10.3168/jds.2015-9914.

Dórea, J. R. R., G. J. M. Rosa, K. A. Weld, and L. E. Armentano. 2018. Mining data from milk infrared spectroscopy to improve feed intake predictions in lactating dairy cows. J. Dairy Sci. 101:5878–5889. https://doi.org/10.3168/jds.2017-13997.

Ehret, A., D. Hochstuhl, D. Gianola, and G. Thaller. 2015. Application of neural networks with back-propagation to genome-enabled prediction of complex traits in Holstein-Friesian and German Fleckvieh cattle. Genet. Sel. Evol. 47:22. https://doi.org/10.1186/s12711-015-0097-5.

Felipe, V. P. S., M. A. Silva, B. D. Valente, and G. J. M. Rosa. 2015. Using multiple regression, Bayesian networks and artificial neural networks for prediction of total egg production in European quails based on earlier expressed phenotypes. Poult. Sci. 94:772–780. https://doi.org/10.3382/ps/pev031.

Feng, N., F. Wang, and Y. Qiu. 2006. Novel approach for promoting the generalization ability of neural networks. Int. J. Signal Process. 2:131–135.

Fleming, A., F. S. Schenkel, J. Chen, F. Malchiodi, V. Bonfatti, R. A. Ali, B. Mallard, M. Corredig, and F. Miglior. 2017. Prediction of milk fatty acid content with mid-infrared spectroscopy in Canadian dairy cattle using differently distributed model development sets. J. Dairy Sci. 100:5073–5081. https://doi.org/10.3168/jds.2016-12102.

Foresee, F. D., and M. T. Hagan. 1997. Gauss-Newton approximation to Bayesian learning. Pages 1930–1935 in Proceedings of International Conference on Neural Networks (ICNN'97). IEEE.

Gianola, D., H. Okut, K. A. Weigel, and G. J. M. Rosa. 2011. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. BMC Genet. 12:87. https://doi.org/10.1186/1471-2156-12-87.

Guo, P., M. R. Lyu, and C. L. P. Chen. 2003. Regularization parameter estimation for feedforward neural networks. IEEE Trans. Syst. Man Cybern. B Cybern. 33:35–44. https://doi.org/10.1109/TSMCB.2003.808176.

Hagan, M. T., and M. B. Menhaj. 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Netw. 5:989–993. https://doi.org/10.1109/72.329697.

Hajmeer, M., I. Basheer, and D. O. Cliver. 2006. Survival curves of Listeria monocytogenes in chorizos modeled with artificial neural

networks. Food Microbiol. 23:561–570. https://doi.org/10.1016/j.fm.2005.09.011.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. Neural networks. Pages 389–416 in The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer.

Haykin, S. 2009. Multilayer perceptions. Pages 123–217 in Neural Networks: A Comprehensive Foundation. 3rd ed. Pearson Prentice Hall.

Hemme, T., M. M. Uddin, and O. A. Ndambi. 2014. Benchmarking cost of milk production in 46 countries. J. Rev. Glob. Econ. 3:254–270. https://doi.org/10.6000/1929-7092.2014.03.20.

Jamrozik, J., and G. J. Kistemaker. 2020. Feed efficiency evaluation: August 2020 test run. Accessed Aug. 11, 2022. http://www.cdn.ca/Articles/GEBOCT2020/FE%20Report%20-%20September%202020.pdf.

Lahart, B., S. McParland, E. Kennedy, T. M. Boland, T. Condon, M. Williams, N. Galvin, B. McCarthy, and F. Buckley. 2019. Predicting the dry matter intake of grazing dairy cows using infrared reflectance spectroscopy analysis. J. Dairy Sci. 102:8907–8918. https://doi.org/10.3168/jds.2019-16363.

Lin, M.-I. B., W. A. Groves, A. Freivalds, E. G. Lee, and M. Harper. 2012. Comparison of artificial neural network (ANN) and partial least squares (PLS) regression models for predicting respiratory ventilation: An exploratory study. Eur. J. Appl. Physiol. 112:1603–1611. https://doi.org/10.1007/s00421-011-2118-6.

MacKay, D. J. C. 1992. Bayesian interpolation. Neural Comput. 4:415–447. https://doi.org/10.1162/neco.1992.4.3.415.

MacKay, D. J. C., and D. J. C. MacKay. 2003. Neural networks. Pages 468–549 in Information Theory, Inference and Learning Algorithms. Cambridge University Press.

Martens, H., and T. Naes. 1989. Multivariate Calibration. John Willey & Sons Inc.

McDermott, A., G. Visentin, M. De Marchi, D. P. Berry, M. A. Fenelon, P. M. O'Connor, O. A. 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., G. Banos, E. Wall, M. P. Coffey, H. Soyeurt, R. F. Veerkamp, and D. P. Berry. 2011. The use of mid-infrared spectrometry to predict body energy status of Holstein cows. J. Dairy Sci. 94:3651–3661.

McParland, S., E. Lewis, E. Kennedy, S. G. Moore, B. McCarthy, M. O'Donovan, S. T. Butler, J. E. 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.

Mevik, B. H., R. Wehrens, and K. H. Liland. 2013. pls: Partial least squares and principal component regression. R Packag. version 2.3-0. Accessed Aug. 11, 2022. http://www2.uaem.mx/r-mirror/web/packages/pls/pls.pdf.

Møller, M. F. 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw. 6:525–533. https://doi.org/10.1016/S0893-6080(05)80056-5.

Muñoz, C., P. A. Letelier, E. M. Ungerfeld, J. M. Morales, S. Hube, and L. A. Pérez-Prieto. 2016. Effects of pregrazing herbage mass in late spring on enteric methane emissions, dry matter intake, and milk production of dairy cows. J. Dairy Sci. 99:7945–7955. https://doi.org/10.3168/jds.2016-10919.

Okut, H., D. Gianola, G. J. M. Rosa, and K. A. Weigel. 2011. Prediction of body mass index in mice using dense molecular markers and a regularized neural network. Genet. Res. (Camb) 93:189–201. https://doi.org/10.1017/S0016672310000662.

Okut, H., X.-L. Wu, G. J. M. Rosa, S. Bauck, B. W. Woodward, R. D. Schnabel, J. F. Taylor, and D. Gianola. 2013. Predicting expected progeny difference for marbling score in Angus cattle using artificial neural networks and Bayesian regression models. Genet. Sel. Evol. 45:34. https://doi.org/10.1186/1297-9686-45-34.

Perai, A. H., H. Nassiri Moghaddam, S. Asadpour, J. Bahrampour, and G. Mansoori. 2010. A comparison of artificial neural networks with other statistical approaches for the prediction of true metabolizable energy of meat and bone meal. Poult. Sci. 89:1562–1568. https://doi.org/10.3382/ps.2010-00639.

Pérez-Rodríguez, P., D. Gianola, K. A. Weigel, G. J. M. Rosa, and J. Crossa. 2013. An R package for fitting Bayesian regularized neural networks with applications in animal breeding. J. Anim. Sci. 91:3522–3531. https://doi.org/10.2527/jas.2012-6162.

Rinnan, A., L. Norgaard, F. van den Berg, J. Thygesen, R. Bro, and S. B. Engelsen. 2009. Data pre-processing. Pages 29–50 in Infrared Spectroscopy for Food Quality Analysis and Control. Academic Press.

Rutten, M. J. M., H. Bovenhuis, J. M. L. Heck, and J. A. M. Van Arendonk. 2011. Predicting bovine milk protein composition based on Fourier transform infrared spectra. J. Dairy Sci. 94:5683–5690. https://doi.org/10.3168/jds.2011-4520.

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.

Seymour, D. J., A. Cánovas, T. C. S. Chud, J. P. Cant, V. R. Osborne, C. F. Baes, F. S. Schenkel, and F. Miglior. 2020. The dynamic behavior of feed efficiency in primiparous dairy cattle. J. Dairy Sci. 103:1528–1540. https://doi.org/10.3168/jds.2019-17414.

Shadpour, S., T. C. S. Chud, D. Hailemariam, H. R. Oliveira, G. Plastow, P. Stothard, J. Lassen, R. Baldwin, F. Miglior, C. F. Baes, D. Tulpan, and F. S. Schenkel. 2022. Supplementary Tables [Data set]. Zenodo. https://doi.org/10.5281/zenodo.6632059.

Shetty, N., P. Løvendahl, M. S. Lund, and A. J. Buitenhuis. 2017. Prediction and validation of residual feed intake and dry matter intake in Danish lactating dairy cows using mid-infrared spectroscopy of milk. J. Dairy Sci. 100:253–264. https://doi.org/10.3168/jds.2016-11609.

Sorich, M. J., J. O. Miners, R. A. McKinnon, D. A. Winkler, F. R. Burden, and P. A. Smith. 2003. Comparison of linear and nonlinear classification algorithms for the prediction of drug and chemical metabolism by human UDP-glucuronosyltransferase isoforms. J. Chem. Inf. Comput. Sci. 43:2019–2024. https://doi.org/10.1021/ci034108k.

Soyeurt, H., D. Bruwier, J.-M. Romnee, N. Gengler, C. Bertozzi, D. Veselko, and P. Dardenne. 2009. Potential estimation of major mineral contents in cow milk using mid-infrared spectrometry. J. Dairy Sci. 92:2444–2454. https://doi.org/10.3168/jds.2008-1734.

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.

Soyeurt, H., F. Dehareng, N. Gengler, S. McParland, E. Wall, D. P. Berry, M. Coffey, and P. Dardenne. 2011. Mid-infrared prediction of bovine milk fatty acids across multiple breeds, production systems, and countries. J. Dairy Sci. 94:1657–1667. https://doi.org/10.3168/jds.2010-3408.

Titterington, D. M. 2004. Bayesian methods for neural networks and related models. Stat. Sci. 19:128–139. https://doi.org/10.1214/088342304000000099.

UN (United Nations). 2019. Growing at a slower pace, world population is expected to reach 9.7 billion in 2050 and could peak at nearly 11 billion around 2100. UN Department of Economic and Social Affairs. Accessed Aug. 25, 2021. https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html.

van der Drift, S. G. A., R. Jorritsma, J. T. Schonewille, H. M. Knijn, and J. A. Stegeman. 2012. Routine detection of hyperketonemia in dairy cows using Fourier transform infrared spectroscopy analysis of β-hydroxybutyrate and acetone in milk in combination with test-day information. J. Dairy Sci. 95:4886–4898. https://doi.org/10.3168/jds.2011-4417.

van Knegsel, A. T. M., S. G. A. van der Drift, M. Horneman, A. P. W. de Roos, B. Kemp, and E. A. M. Graat. 2010. Ketone body concentration in milk determined by Fourier transform infrared spectroscopy: Value for the detection of hyperketonemia in dairy cows. J. Dairy Sci. 93:3065–3069. https://doi.org/10.3168/jds.2009-2847.

Veerkamp, R. F., and S. Brotherstone. 1997. Genetic correlations between linear type traits, food intake, live weight and condition score in Holstein Friesian dairy cattle. Anim. Sci. 64:385–392. https://doi.org/10.1017/S1357729800015976.

Wallén, S. E., M. Lillehammer, and T. H. E. Meuwissen. 2017. Strategies for implementing genomic selection for feed efficiency in dairy cattle breeding schemes. J. Dairy Sci. 100:6327–6336. https://doi.org/10.3168/jds.2016-11458.

Wallén, S. E., E. Prestløkken, T. H. E. Meuwissen, S. McParland, and D. P. Berry. 2018. Milk mid-infrared spectral data as a tool to predict feed intake in lactating Norwegian Red dairy cows. J. Dairy Sci. 101:6232–6243. https://doi.org/10.3168/jds.2017-13874.

Wang, H. J., F. Ji, C. S. Leung, and P. F. Sum. 2009. Regularization parameter selection for faulty neural networks. Int. J. Intell. Syst. Technol. 4:45–48.

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.

ORCIDS

Saeed Shadpour https://orcid.org/0000-0002-2842-031X
Tatiane C. S. Chud https://orcid.org/0000-0001-7559-1165
Dagnachew Hailemariam https://orcid.org/0000-0002-0030-7729
Hinayah R. Oliveira https://orcid.org/0000-0002-0355-8902
Graham Plastow https://orcid.org/0000-0002-3774-3110
Paul Stothard https://orcid.org/0000-0003-4263-969X
Jan Lassen https://orcid.org/0000-0002-1338-8644
Ransom Baldwin https://orcid.org/0000-0002-0753-4377
Filippo Miglior https://orcid.org/0000-0003-2345-8842
Christine F. Baes https://orcid.org/0000-0001-6614-8890
Dan Tulpan https://orcid.org/0000-0003-1100-646X
Flavio S. Schenkel https://orcid.org/0000-0001-8700-0633