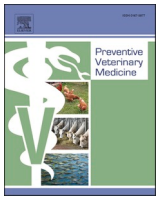
Preventive Veterinary Medicine 193 (2021) 105422

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

journal homepage: www.elsevier.com/locate/prevetmed



Full model selection using regression trees for numeric predictions of biomarkers for metabolic challenges in dairy cows

J.F. Mandujano Reyes [a,*], E. Walleser [a], S. Hachenberg [b], S. Gruber [c], M. Kammer [d], C. Baumgartner [e], R. Mansfeld [c], K. Anklam [a], D. Döpfer [a]

[a] Department of Medical Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI, USA
[b] German Association for Performance and Quality Testing e.V. (DLQ), 53111, Bonn, Germany
[c] Clinic for Ruminants with Ambulatory and Herd Health Services, Ludwig-Maximilian-University Munich (LMU Munich), 85764, Oberschleißheim, Germany
[d] LKV Bayern e. V. (Dairy Herd Improvement Association of Bavaria), 80687, Munich, Germany
[e] Bavarian Association for Raw Milk Testing e. V. (MPR), 85283, Wolnzach, Germany

ARTICLE INFO

Keywords:
Metabolic challenge
Fourier-transform infrared spectroscopy
Full model selection
Machine learning
Dairy cows

ABSTRACT

Dairy cows suffer poor metabolic adaptation syndrome (PMAS)[1] during early post-calving periods caused by negative energy balance. Measurement of blood beta-hydroxy butyric acid (BHBA)[2] and blood non-esterified fatty acids (NEFA)[3] allow early and accurate detection of negative energy balance. Machine learning prediction of blood BHBA and blood NEFA using milk testing samples represents an opportunity to identify at-risk animals, using less labor than direct blood testing methods. Routine milk testing on modern dairies and computer record keeping provide an immense amount of data which can then be used in machine learning models. Previous research for predicting blood metabolites using Fourier-transform infrared spectroscopy (FTIR)[4] milk data has focused mainly on individual models rather than a comparison among the models. Full model selection is the process of comparing different combinations of pre-processing methods, variable selection, and statistical learning algorithms to determine which model results in the lowest prediction error for a given dataset. For this project we used a full model selection approach with regression trees (rtFMS)[5] . rtFMS uses the cross-validated performance of different model configurations to feed a regression tree for selecting a final model. A total of 384 possible model configurations (algorithms, predictors and data preprocessing options) for each outcome (blood BHBA and blood NEFA) were considered in the rtFMS technique. rtFMS allows direct comparison of multiple modeling approaches reducing bias due to empirical knowledge, modeling habits, or preferences, identifying the model with minimal root mean squared prediction error (RMSE)[6] . An elastic net regression model was selected as the best performing model for both biomarkers. The input data for blood BHBA predictions were FTIR milk spectra, with a second derivative pre-processing, and a filter with 212 wave numbers, obtaining RMSE = 0.354 (0.328−0.392). The best performing model for blood NEFA had input data of FTIR milk spectra, with a second derivative pre-processing, and a filter with 212 wave numbers filter along with the time of milking, obtaining RMSE = 0.601 (0.564−0.654). The comparison of multiple modeling strategies, conducted by rtFMS, present an option for improved FTIR prediction models of blood BHBA and blood NEFA by reducing error due to human bias. The implementation of rtFMS to design future prediction models can guide model inputs and features. Our prediction models have the potential to increase early detection of metabolic disorders in dairy cows during the transition period.

* Corresponding author at: School of Veterinary Medicine, University of Wisconsin-Madison, 2015 Linden Dr, Madison, WI, 53706-1100, USA.
E-mail address: mandujanorey@wisc.edu (J.F. Mandujano Reyes).

[1] PMAS: Poor metabolic adaptation syndrome.
[2] BHBA: beta-hydroxy butyric acid.
[3] NEFA: non-esterified fatty acids.
[4] FTIR: Fourier-transform infrared spectroscopy.
[5] rtFMS: full model selection approach with regression trees.
[6] RMSE: root mean squared prediction error.

https://doi.org/10.1016/j.prevetmed.2021.105422
Received 9 February 2021; Received in revised form 19 May 2021; Accepted 28 June 2021
Available online 30 June 2021

1. Introduction

Dairy cows suffering from metabolic disorders related to negative energy balance are at an increased risk of disease and reduced milk production (Duffield et al., 2009). Management of metabolic disturbances is routinely accomplished on farm through blood testing for beta-hydroxy butyric acid (BHBA) and laboratory tests for blood non-esterified fatty acids (NEFA) (Oetzel, 2004). Alternative strategies that minimize on-farm labor and provide accurate results would likely be adopted by the dairy industry. The use of Fourier-transform infrared spectroscopy (FTIR) data from milk test samples, with additional production information, is one strategy to use routine milk testing samples instead of invasive blood tests (Grelet et al., 2016; Hansen, 1999; Heuer et al., 2001; Tremblay et al., 2019).

Prediction models for blood BHBA and NEFA using milk testing data is an automated, early, and reduced labor approach to metabolic disease detection. Different pre-processing techniques for data, as well as different statistical algorithms for prediction modeling are available. Most models for prediction of metabolic status using FTIR data as input have been performed with partial least squares regression (Pralle et al., 2018; Rohman et al., 2014; Sim and Jeffrey Kimura, 2019). Other studies have been conducted using different approaches including multiple linear regression and artificial neural networks (Pralle et al., 2018). Due to the complexity of the problem and the multiple options available, full model selection (FMS) can be used to identify the best performing prediction model when numerous different options for modeling should be compared (Díaz-Pacheco et al., 2018). The FMS technique is the selection process for the combination of pre-processing methods, feature selection, and statistical learning algorithms that results in the lowest prediction error (or highest prediction accuracy) for a given data set (Escalante et al., 2009). The regression tree full model selection (rtFMS) approach was previously reported as an application of FMS to identify the best performing prediction models for classification of elevated blood BHBA and blood NEFA using milk FTIR and cow information (Gruber et al., 2021; Tremblay et al., 2019). Our objective was to implement rtFMS for numeric predictions of blood BHBA and blood NEFA. rtFMS is considered to be a non-parametric and systematic approach towards reducing user bias in terms of modeling and pre-processing choices in the face of numerous options and habits for such choices (Gruber et al., 2021; Tremblay et al., 2019). With the rtFMS methodology, the evaluation of the hypothesis of no association between the outcome variable (root mean squared error, (RMSE)) and the options for predictive model using multiple hypothesis testing is performed (Hothorn et al., 2006).

To permit a user adaptable implementation of blood BHBA and blood NEFA prediction models it is of importance to consider numeric concentration values as final output. Numeric concentration predictions may have greater utility than direct classification models. Different cut off values for blood NEFA have been reported, such as <0.39 mmol/L (Giuliodori et al., 2011) and ≥ 0.7 mmol/L (Tremblay et al., 2018) to identify low and high Poor Metabolic Adaptation Syndrome (PMAS). Similarly, different values for blood BHBA have been used, for example >1.2 mmol/L (Iwersen et al., 2013) and >1.4 mmol/L (Carrier et al., 2004) for subclinical ketosis. Variable choices of cut-off values for both biomarkers, for disease detection indicate that predicting numerical values for blood BHBA and blood NEFA would allow users to use different thresholds and potentially change treatments based on blood levels.

The objective of this study was to use routinely obtained milk testing data, including milk FTIR information and cow information to identify the best performing prediction models for concentrations of blood BHBA and blood NEFA among a wide range of combination of pre-processing methods, feature selection, and statistical learning algorithms by using rtFMS. By predicting numerical BHBA and NEFA concentrations accurately we may improve flexibility of model usage and improve user decisions. At the same time, we aim to provide a numerical prediction of these biomarkers, in order to detect PMAS, not depending on any cut-off value using the rtFMS approach to achieve this. An additional objective of this study was to evaluate the performance of the rtFMS approach for numerical predictions comparing the efficiency of our models with similar studies.

2. Materials and methods

2.1. The data set

The data set used in this study (Q Check data set) was collected weekly on ten dairy farms in the states of Bavaria and Thuringia in Germany from January 2018 until December 2018. Among the farms, there were eight with automated milking parlors and two farms with conventional milking parlors. The data set included the following cow information: days in milk (DIM), parity (coded as primiparous vs multiparous), milk production per day in kg (milk yield), breed of the cows, (Simmental or Holstein-Friesian), time point of milking per sample (AM or PM), milk Fourier-transform infrared spectroscopy (FTIR) data, FOSS milk fatty acid panel predictions (FOSS, Hilleroed, Denmark), and blood measurements for concentration of BHBA and NEFA in mmol/L. A log transformation for the blood BHBA and NEFA values was used to normalize the skewed distributions to accommodate the assumptions of the linear regression-based algorithms considered (therefore, if noting else is mentioned, the showed RMSE will be in logarithmic scale).

The Q Check data set contained 11,494 observation points from 2,456 cows (2,058 Holstein, 398 Simmental) of which 1,476 were multiparous and 980 primiparous. After selecting only observations between 5 and 50 DIM, and removing observations without complete fatty acid panels, 9,660 of the 11,494 observations remained in the data set. A subset of 9,442 observations had time of milking being 4,634 in the morning (AM) and 4,808 in the evening (PM). Table 1 presents descriptive statistics of DIM, milk yield, and blood BHBA (Median = 0.70 mmol/L, Interquartile interval = 0.55−0.86) and NEFA (Median = 0.20 mmol/L, Interquartile interval = 0.11−0.37) concentrations.

2.2. Options for the model

Three groups of options for the model were considered before the data analysis: (i) the types of input variables (fatty acid panels or milk FTIR data), inclusion of DIM, parity (heifer or cow), milk yield in kg, breed (Holstein or Simmental), and whether or not using the time of milk sampling (AM or PM); (ii) the type of pre-processing of FTIR data (standardization as reported by Grelet et al. (2016) or FTIR data processed using first or second derivative of the values) and filtering for 212 informative wave numbers (EMR212), of 1060 original wave numbers, for FTIR that overlap between FOSS and Bentley instruments (FOSS, Hilleroed, Denmark; Bentley Instruments Inc. Chaska, MN, USA), and (iii) selection of four Machine Learning (ML) algorithms related to linear regression models (specified below).

Table 1
Descriptive statistics of Q Check data set: days in milk, milk yield, and blood BHBA and NEFA concentrations (n = 9,960).

Variable	1st Quantile	Median	Mean	3rd Quantile
Days in Milk	14.00	22.00	22.85	30.00
Milk Yield (Kg)	27.40	33.10	33.48	39.50
Blood NEFA[a] concentration (mmol/L)	0.11	0.20	0.29	0.37
Blood BHBA[b] concentration (mmol/L)	0.55	0.70	0.75	0.86

[a] NEFA: non-Esterified Fatty Acids.
[b] BHBA: Beta-Hydroxy Butyric Acid.

As shown in Table 2, the milk fatty acid data was organized in panels. The FOSS fatty acids panel FAF1 contains: saturated, monounsaturated, and polyunsaturated fatty acids; the FOSS fatty acids panel FAF2 included: short chain (C4, C6, C8, C10), medium chain (C12, C14, C16), long chain (C18), and the major fatty acids (C14:0, C16:0, C18:0, C18:1); and the FOSS fatty acids panel FAF3 contains: de novo ($\leq$ C14), mixed (C16) and performed fatty acids (C15, C17, $\geq$ C18). Furthermore, the panel FAF is the combination of all the panels FAF1, FAF2 and FAF3.

The aforementioned standardization of FTIR data (coded as EMR standardization) was conducted according to the OptiMIR project (Grelet et al., 2016). All numeric derivatives were performed using difference quotients $\frac{f(x_2)-f(x_1)}{x_2-x_1}$, where x represents the wavenumber and f (x) is the observed value for the wavenumber x. Therefore, regarding derivative based pre-processing: in addition to the first and second derivative with gap (the spacing between points x_2 and x_1) equal to one (FD, SD), a first and second derivatives using a gap of four were considered. Since the derivatives with gap equal 4 are computed according to the OptiMIR project recommendations, these are the called as EMR derivatives (FD.EMR, SD.EMR).

We selected four ML algorithms to evaluate based on previous published research with milk FTIR predictions: (i) The method "glmnet" for elastic net regression (Tremblay et al., 2019), (ii) "earth" for multivariate adaptative regression splines (Tremblay et al., 2019), (iii) "pls" for partial least square regression (Paradkar and Irudayaraj, 2002), (iv) "pcr" for principal component regression (Paradkar and Irudayaraj, 2002), from the caret package (Kuhn, 2008) in R software version 3.6.3 (R Core Team, 2020) were used. Table 2 shows the options for the prediction models.

Table 2
Options for comparison using regression tree full model selection to build prediction models for two biomarkers of metabolic challenges in dairy cows: non-esterified fatty acids (NEFA) and beta-hydroxy butyric acid (BHBA).

Area	Category	Options
Input subset	Milk Data	Fourier transform infrared spectroscopy FTIR (IR)
		FOSS milk fatty acid panels (FAF1)[a]
		FOSS milk fatty acid panels (FAF2)[b]
		FOSS milk fatty acid panels (FAF3)[c]
		FOSS milk fatty acid panels (FAF)[d]
	Breed Information	Use of breed as categorical variable (Breed.Yes)
		Exclusion of breed as categorical variable (Breed.No)
	Time of Milk Sampling	Use of time (AM or PM) when the cow was milked (Milk.Model.Yes)
		Exclusion of time when the cow was milked (Milk.Model.No)
Pre-processing and Filtering	Standardization EMR[e]	Use of standardization EMR (EMR.STAND)
		Exclusion of standardization EMR (EMR.NONE)
	Derivative Pre-processing	Exclusion of derivative based standardization (Raw)
		First derivative (FD)
		Second derivative (SD)
		First derivative according to EMR (FD.EMR)[e]
		Second derivative according to EMR (SD.EMR)[e]
	Filter[f]	Use of 212 wave numbers EMR filter (EMR212)[f]
		Exclusion of any filter (AllWN)
Algorithm	Regression Model	Elastic Net Regression (GLMNET)
		Multivariate Adaptive Regression Splines (MARS)
		Partial Least Squares Regression (PLSReg)
		Principal Component Regression (PCR)

[a] FAF1: saturated, monounsaturated and polyunsaturated fatty acids.
[b] FAF2: short chain (C4, C6, C8, C10), medium chain (C12, C14, C16), long chain (C18) and major fatty acids (C14:0, C16:0, C18:0, C18:1).
[c] FAF3: de novo ($\leq$C14), mixed (C16) and performed fatty acids (C15, C17, $\geq$ C18).
[d] FAF: combination of all the panels FAF1, FAF2 and FAF3.
[e] First and second derivatives using a gap of four.
[f] The 212 informative wave numbers for FTIR that overlap between FOSS and Bentley machines (Grelet et al., 2016).

2.3. Model training

Combinations of input variables, pre-processing, and ML algorithms allowed us to evaluate the prediction performances of the respective models. Since the pre-processing and filtering options (for FTIR data) in Table 2 are not applicable to the fatty acid panels, a total of 384 models were evaluated for blood BHBA and blood NEFA. All hyper-parameters in the considered models were automatically selected using a grid search comparing 10 meaningful values for each hyper-parameter. This was accomplished using the "tuneLength" parameter in caret package (Kuhn and Johnson, 2013).

Using the root mean squared error (RMSE) as a performance metric for the models, 10 repeated iterations of 10−fold cross-validation were conducted by applying the "repeatedcv" method (specifying repeats n = 10, and fold number k = 10) in the function "trainControl" that feeds the "trControl" argument in the function "train" of caret package (Kuhn, 2008). In general, the n iterated k-fold cross-validation works by randomly partitioning the complete data set into k groups (called folds). Then, the model is trained using k-1 folds and holding one fold out as test data set to validate the model. This step is carried out using each of the k folds once as test data. Then, the process is repeated n times always using a different random partition of the data. In the end, all of the observations will have been used once in a validation data set (Kuhn and Johnson, 2013; Tremblay et al., 2019). The function "groupKFold" was used to customize the folds in the 10 iterated 10-fold cross-validation process by splitting the observations into training and test data subsets separating by farm for each cross-validation iteration. This approach reduces bias that may be incorporated into the model when including observations of cows from the same farm in both the training and test data sets.

2.4. Regression tree

Finally, rtFMS was used to select the model that minimizes the prediction error (RMSE). The function "ctree" from partykit package (Hothorn and Zeileis, 2015) was used to perform the regression tree. Eq. (1) describes the regression tree that has RMSE as outcome variable:

$$RMSE \sim Input + Breed + Time + EMR\ Standardization + Derivative + 212\ Filter + ML\ Algorithm. \quad (1)$$

In Eq. (1), Input = the choice of input variables; Breed = Holstein compared to Simmental cows; Time = AM versus PM; EMR Standardization = the approach for standardizing the IR spectral data as reported by Grelet et al. (2016); Derivative = first or second derivative for the IR spectral data (with gap equal one or equal four); 212 Filter = the subset of 212 wavenumbers that are contained in the FOSS and Bentley spectral data (binary decision); and ML Algorithm = the choice of GLMNET, MARS, PLSReg, or PCR methods (Table 2).

The rtFMS approach makes use of conditional regression trees to split the model options shown in Table 2 into groups of modeling choices (coded as predictors) according to the statistical significance of their association with the outcome variable RMSE. Testing the null hypothesis of independence between the model options and RMSE response, along with stopping criteria at level $\alpha = 0.05$ for the branching in the resulting tree, accomplishes the optimized split of models into groups

characterized by different prediction performances (Hothorn et al., 2006). Using the branching decisions of the regression trees, the models were partitioned into nodes that maximize differences in their prediction performance. This process is repeated for each node until the null hypothesis of independence between the response and the covariates cannot be rejected at a pre-specified level α (Hothorn et al., 2006). The level $\alpha = 0.05$ was used in this study, and the presented *P*-values are Bonferroni corrected for multiple testing (Hothorn et al., 2006). This process ends with the selection of a set of models that are expected to have a similar performance among which we can select a final model for each biomarker, blood BHBA and NEFA.

2.5. External validation

In order to compute the performance of the best models, we selected an independent data set reported and described by Tremblay et al. (2018), being the FSM-IRMi data set. Briefly, the data consist of 1,034 observations, collected on farms in Bavaria, Germany. The data set included data from 26 farms, 346 cows and 115 sampling days (from 5 to 50 DIM period). Blood values for BHBA and NEFA, cow and production information, in addition to milk FTIR data were available. Using these data points, the RMSE of the predictions, as well as bootstrap confidence intervals, were computed.

3. Results

The rtFMS approach requires to evaluate the performance of every combination of predictive modeling configurations. Looping through all the possible modeling options under consideration, 384 models for each of the two biomarkers were fitted and cross-validated. The computational time was approximately 1 week (128 GB ram memory, 24 cores 2.10 GHz).

3.1. Blood BHBA

The 384 models for BHBA had a mean RMSE of 0.399 (SD = 0.029). In addition, the rtFMS algorithm had 18 terminal nodes. Fig. 1 illustrates the branch with minimal RMSE was the 3rd from the left that summarizes ten models with median RMSE of 0.359 (SD = 0.003). All nodes and branches of the complete regression tree are shown in Fig. 1. The best performing branch of the models was selected after 6 decision nodes: (1) model algorithm GLMNET, MARS ($P < 0.001$), (2) input FAF, IR ($P < 0.001$), (3) exclusion of Breed ($P < 0.001$), (4) use of filter EMR212 ($P < 0.001$, since EMR212 is a filtering option for IR data only, the FAF panel is not part of the best branch), (5) model GLMNET ($P < 0.001$), (6) exclusion of time of Milking ($P = 0.014$).

After examining the ten models with similar performance from the 3rd branch of the regression tree, one final prediction model was chosen. The final model with minimum RMSE amongst the models in the best performing branch was characterized by FTIR spectral data (IR), second derivative of EMR (SD.EMR), no EMR standardization (EMR.NONE), no breed information (Breed.No), no Time of Milking (Milk.Model.No), and using the 212 wave number filter (EMR212). Using the data set from the final prediction model we fit an Elastic Net (GLMNET) with optimal values for $\alpha = 0.2$ and $\lambda = 8.84 \times 10^{-5}$. Obtaining an RMSE = 0.354 (0.328−0.392), Coefficient of Determination (R^2) = 0.363 (0.277−0.454) and Mean Absolute Error (MAE) = 0.280 (0.255−0.319).

Fig. 2 illustrates that the final prediction model performs well for

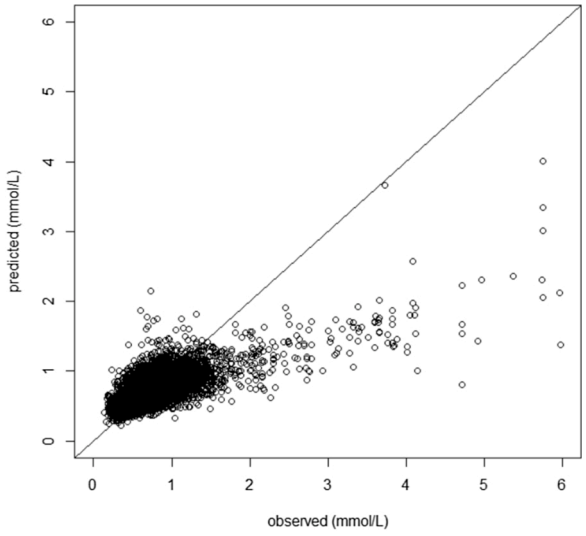


Fig. 2. Observed versus predicted beta-hydroxy butyric acid (BHBA) concentration in blood. The black line represents the relation 1:1; (n = 9,660).

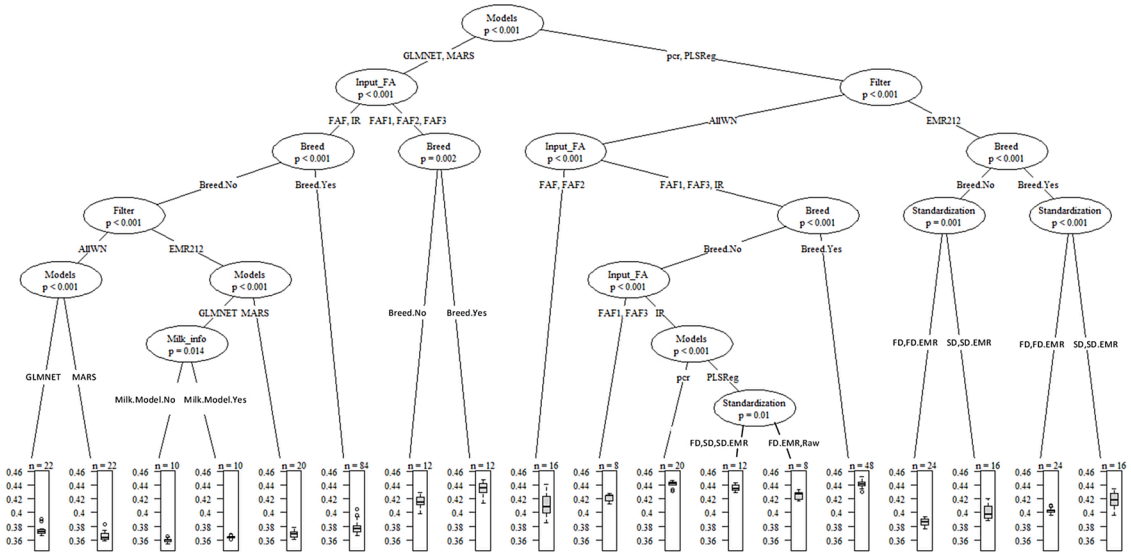


Fig. 1. Full model selection with regression trees (rtFMS) for blood beta-hydroxy butyric acid (BHBA). The root mean squared error of the 384 BHBA predictions models were fed into the regression tree; n = number of models in each branch; boxplots represent the root mean squared error of models per terminal branch. The bottom and top of the box represent the 25th and 75th percentiles, respectively, and the horizontal line inside the box is the median.

blood BHBA values between 0 and 2 mmol/L, but tends to underestimate larger values of blood BHBA. Table 3 presents the top 20 key predictors for blood BHBA from the final model computed with function "varImp" from the caret library. All the key predictors were wave numbers from the FTIR data.

3.2. Blood NEFA

For NEFA, the 384 models had a mean RMSE of 0.697 (SD = 0.100). In this case, the rtFMS algorithm had 21 terminal nodes. As shown in Fig. 3, the branch with minimal RMSE was the 3rd from the left, with 32 models in the group, and a median RMSE of 0.612 (SD = 0.005). The final model was selected after five decision nodes: (1) Model algorithms GLMNET, MARS ($P < 0.001$), (2) input FAF, FAF1, IR ($P < 0.001$), (3) derivative based standardization FD, FD.EMR, SD, SD.EMR ($P < 0.001$, since derivatives are preprocessing options for IR data only, FAF and FAF1 panels are no longer in the best branch), (4) exclusion of breed ($P < 0.001$), (5) filter EMR212 ($P < 0.001$). Fig. 3 shows the complete tree with all nodes and decisions regarding the model options.

The best branch of the tree (3rd from the left in Fig. 3) have 32 models with similar statistical performance, we choose one (the one with smallest RMSE) for further evaluation. The data for the final model were using raw FTIR spectral data (IR), applying second derivative (SD), excluding the EMR standardization (EMR.NONE), excluding breed information (Breed.No), using Time of Milking (Milk.Model.Yes), and using the 212 wave number filter (EMR212). Using the data set from the final prediction model we fit an Elastic Net (GLMNET), with optimal values for $\alpha = 0.2$ and $\lambda = 1.943 \times 10^{-4}$. In this way, an RMSE = 0.601 (0.564−0.654), $R^2 = 0.466$ (0.395−0.525) and MAE = 0.481 (0.449−0.528) were obtained.

Fig. 4 shows good performance of the model for values of NEFA between 0 and 1.5, and a discrepancy in both directions between predicted and observed for higher values. Table 4 presents the top 20 key predictors, computed using the function "varImp" from caret, from the selected final model for NEFA. As with the BHBA, all the key predictors were wave numbers from FTIR data for NEFA.

Table 3
The 20 most important predictors for the best performing prediction model for Beta-Hydroxy Butyric Acid (BHBA) chosen using rtFMS. Model: Elastic Net (GLMNET) with FTIR spectral data (IR), second derivative of EMR (SD. EMR), excluding EMR standardization (EMR.NONE), excluding breed information (Breed.No), excluding Time of Milking (Milk.Model.No), and using the 212 wave number filter (EMR212); (n = 9,660).

Predictor[1] (cm^{-1})	Importance[2]
1804.14	100
1781.01	91.047
1784.865	74.337
1788.72	64.689
1796.43	64.541
1210.47	47.852
1241.31	43.501
1807.995	39.01
1268.295	29.029
1264.44	27.688
1260.585	27.447
1129.515	26.876
1287.57	25.989
1256.73	25.42
967.605	25.103
1102.53	19.878
1345.395	19.556
1245.165	19.355
1252.875	17.975
1295.28	16.896

[1] Wavenumbers for IR.
[2] Scaled to 100.

3.3. External validation

Given a lack of time of milking information in the FSM-IRMi validation data set, we excluded that variable in the blood NEFA model. Nonetheless, the best performing model for blood NEFA excluding the variable "time of milking" was included in the best performing branch of models from the rtFMS. Therefore, regarding blood NEFA predictions, we expect a similar performance for the FSM-IRMi as obtained with the chosen model for blood NEFA in the Q Check data set. The RMSE of the predictions of numeric values for blood BHBA and NEFA along with bootstrap confidence intervals were 0.44 (0.42−0.47) and 0.87 (0.83−0.91), respectively.

4. Discussion

4.1. Regression tree full model selection

Machine learning approaches for prediction of blood BHBA and blood NEFA values allow for early and automated detection of cows who may require medical attention post-calving, as a result of metabolic disorders. Enhanced detection methods would likely reduce negative effects correlated with elevated NEFA and BHBA in blood, such as reduced milk production, and increased culling (Chapinal et al., 2012; Seifi et al., 2011).

Given the availability of FTIR and production data from routine milk testing, in combination with ML technologies, it is important to consider all predictive options to implement the best possible model using approaches like FMS. Multiple input and pre-processing options represent a challenge to optimizing prediction models. Moreover, user bias due to empirical knowledge, modeling habits, or preferences for statistical algorithms impact the performance of prediction models. The rtFMS approach provides a systematic way to address these issues by comparing the prediction performance of multiple models reducing user bias at the same time. Other approaches for FMS use different techniques to achieve this goal, an example for another FMS approach is particle swarm optimization (Escalante et al., 2009), but the black box approach of particle swarm optimization was difficult to justify compared to the simpler rtFMS approach. While, in most of the cases, interpretability is preferred in biological applications, rtFMS is seen as a prominent FMS approach.

The rtFMS modeling framework was used in this study for minimizing the RMSE of predicting numerical blood BHBA and blood NEFA. This approach allows more flexibility in model implementation, even when used with binary thresholds. Ketosis treatment varies based on severity and accurate numeric predictions may improve treatment recommendations (Gordon et al., 2013). For these reasons an accurate numeric prediction model is preferable to a classification model as developed in Tremblay et al. (2019).

4.2. Statistical features of selected models

The selected best performing model algorithm for both applications in this study was an Elastic Net, a type of regression model that combines two techniques of penalization: least absolute shrinkage and selection operator (Lasso) and Ridge regression. Therefore, the Elastic Net was meaningful for variable reduction and parameter shrinkage in the face of highly correlated variables, which were a common situation while modeling FTIR data (Zou and Hastie, 2005). Then, two parameters, α and λ, were the hyper-parameters optimized for the Elastic Net model during the caret training process (Friedman et al., 2010; Kuhn and Johnson, 2013) resulting in a prediction model under the constraints of parameter shrinkage and variable reduction. Furthermore, the Elastic Net provides and added benefit not captured in the rtFMS selection approach, unlike the classic PLS regression for FTIR data, variable importance remains interpretable. If both, PLS and Elastic Net models, have the same performance this may be a reason to select Elastic Net

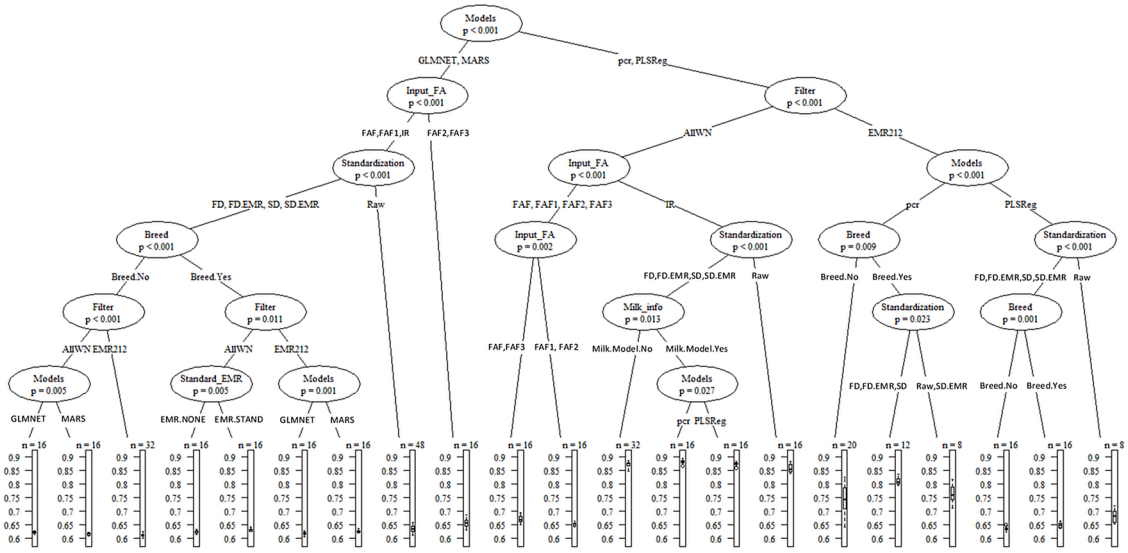


Fig. 3. Full model selection with regression trees (rtFMS) for non-esterified fatty acids (NEFA). The root mean squared error of the 384 NEFA predictions models were fed into the regression tree; n = number of models in each branch; boxplots represent the root mean squared error of models per terminal branch. The bottom and top of the box represent the 25th and 75th percentiles, respectively, and the horizontal line inside the box is the median.

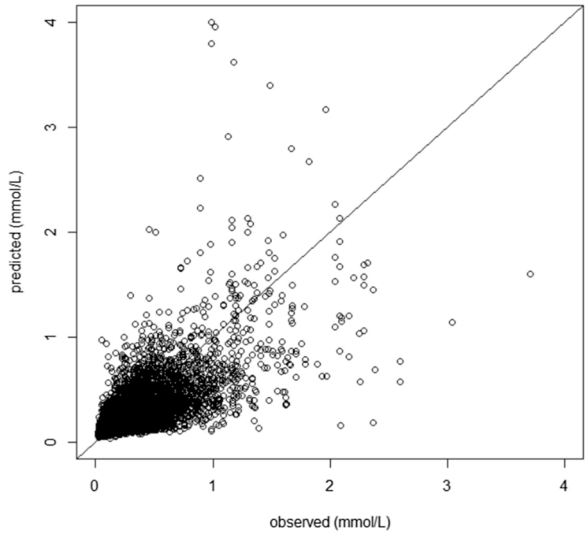


Fig. 4. Observed versus predicted non-esterified fatty acids (NEFA) concentration in blood. The black line represents the relation 1:1; (n = 9,442).

regression over PLS.

During model training of each statistical algorithm, shown in Table 2, 10 repeated iterations of 10−fold cross-validation were used as a validation process. This resampling method can be applied to generate appropriate estimations of the model performance using the complete data set. Furthermore, this technique often produces better performance estimates compared to using a single test data set, because many alternative versions of the data are used (Chollet, 2018; Kuhn and Johnson, 2013). We are aware of the fact that a potential bias can occur when estimating model performance while parameter tuning is performed at the same time. This optimization bias, also called cross validation bias, is most likely insubstantial for moderately large data sets like the one used in this study (Chollet, 2018; Kuhn and Johnson, 2013). This is supported by the fact that k−fold cross-validation is an important validation approach when selecting an optimal model as reported by Krstajic et al. (2014).

Table 4
The 20 most important predictors for the best performing prediction model for non-Esterified Fatty Acids (NEFA) chosen using rtFMS. Model: Elastic net (GLMNET) with FTIR spectral data (IR), second derivative (SD), excluding EMR standardization (EMR.NONE), excluding breed information (Breed.No), using Time of Milking (Milk.Model.Yes), and using the 212 wave number filter (EMR212); (n = 9,442).

Predictor[1] (cm^{-1})	Importance[2]
1796.43	100
1800.285	94.632
1349.25	48.041
1148.79	36.369
1302.99	26.277
1071.69	24.238
1387.8	23.437
1364.67	19.941
2821.86	19.323
1233.6	18.768
1121.805	17.652
1414.785	17.331
1314.555	17.045
1060.125	16.635
998.445	16.247
1368.525	15.842
1391.655	15.214
1229.745	14.948
1052.415	14.856
1094.82	14.065
1175.775	14.061

[1] Wavenumbers for IR.
[2] Scaled to 100.

4.3. Selected modeling options

The information related to DIM, heifer or cow, and milk yield was included in all prediction models. These variables were hypothesized to be important predictors for the two biomarkers, blood BHBA and NEFA. Nevertheless, when the top 20 key predictors were selected, all of them belonged to FTIR data as shown in Tables 3 and 4. Ten of the key predictors for both biomarkers were wavenumbers belonging to 1450−1200 cm^{-1} which is located in the acetone region of milk FTIR data (Heuer et al., 2001). In addition, the top two key predictors for

blood BHBA and NEFA are near 1800 cm-1 in the milk spectrum, which is a region associated to major absorption from fat (Hansen, 1999). This supports the idea that fat metabolism is challenged during the early post-calving period when negative energy balance occurs.

Although the use of fatty acid panels would make physiological sense to predict blood NEFA and blood BHBA, the final prediction models performed better using FTIR data. One explanation is that all FOSS fatty acid panels are actually predictions computed up to 7 days after the milk sampling using the FTIR milk data. This delay in FTIR measurements could result in lower accuracy of the predicted fatty acid panels. Additionally, the fatty acid predictions have an inherent error incorporated by the modeling process. Thus, the use of the predicted fatty acid panels could result in the propagation of prediction error into the final blood NEFA and blood BHBA predictions.

It was hypothesized that Breed would be an important predictor for the numeric prediction models, because the dual-purpose breeds like Simmental may have different metabolism and susceptibility to hyperketonemia compared to Holstein cows (Benedet et al., 2020). However, we did not find a significant association between Breed and either outcome variable. Therefore, breed information can be excluded from the predictors. This would simplify the application of such prediction models for routine milk testing data in the future, particularly in cases where breed is either not recorded or unknown to the producer.

The samples were taken across all times of the day and any time effect of sampling was averaged out in this way. Although not significantly associated with the branching of the rtFMS tree for NEFA and BHBA, Milking time was found to be associated with the outcome in the best performing prediction model for NEFA. The same was reported as well by Mehtiö et al. (2018). Milking time was not a key predictor for the final BHBA prediction model. Therefore, Milking time can likely be ignored based on the results of our models.

For data preprocessing and filtering, the EMR212 subset of wave numbers was chosen for the best performing branch during the rtFMS. This results in the conclusion that one can translate the IR measurements between brands of machines, since the EMR212 subset of wave numbers is present in Bentley machines as well as in FOSS machines (Grelet et al., 2016). In addition, the second derivatives were important preprocessing methods for both outcomes. This type of standardization is useful for removing baseline offset and separation of overlapping peaks (Rieppo et al., 2012; Tremblay et al., 2019). These two techniques, wave number selections and derivatives applied to FTIR measurements are practical solutions for the comparison of milk IR spectral data and data generated by different machines (Tiplady et al., 2019).

In contrast to the objective of standardization among different spectral machines, which is the aim of using EMR standardization, the use of this standardized technique for FTIR measurements was not a significantly associated choice for the variables of the best performing model. This can be explained by the fact that the data were generated on FOSS instruments only.

4.4. Prediction of blood BHBA and blood NEFA

Our prediction models performed well for the prediction of the most common values of blood BHBA and NEFA, but the model predictions for higher values were more variable. This could be due to the fact that the higher values (BHBA $>$ 2 mmol/L and NEFA $>$ 1.5 mmol/L) had limited appearances in the data set and if the values appeared, then their variation was higher than the rest of the observed data. The BHBA $>$ 2 mmol/L represents 1.5 % of the observations; for NEFA 0.8 % of the observations were $>$1.5 mmol/L. Therefore, we recommend to cap the prediction models for BHBA values above 2 mmol/L and for NEFA above 1.5 mmol/L or to interpret the predicted values above these limit values with caution.

4.5. External validation

By employing our model to predict on an external data set we exhibited the generalizability of our models. The FSM-Irmi dataset was suitable for the preprocessing of the best performing prediction models, for both biomarkers, with the consideration of excluding time of milking in the NEFA model. Obtaining RMSE = 0.44 (0.42−0.47) and RMSE = 0.87 (0.83−0.91) for blood BHBA and NEFA respectively, a slightly lower performance than the expected was observed. It is important to note that the error reported is computed using the logarithmic transformation of BHBA and NEFA values. Moreover, because most of the BHBA and NEFA values are near 0 and the fact that the function $f(x) = \log(x)$ tends to $-\infty$ when x tends to 0, a difference of the RMSE in logarithmic scale seems more dramatic compared to only marginal changes on the original scale. Computing the error in the original scale for the FSM-Irmi data set we obtained 0.42 (0.39−0.46) for BHBA and 0.40 (0.37−0.45) for NEFA. Slightly lower prediction performance on external data indicates our prediction model is sensitive to changes in population.

While other numeric prediction models for blood BHBA and blood NEFA have been developed before for smaller data sets (Mehtiö et al., 2018; Pralle et al., 2018), a larger data set was used in this study which could result in including more variation. Additionally, differences among time between milking and blood sampling, different herd sizes, and different milking systems, can have consequences for the performance of our models. Nevertheless, the current findings help to improve and develop future models for more varied and larger data sets.

5. Conclusion

In conclusion, we evaluated multiple modeling options that can be used to predict blood BHBA and blood NEFA from FTIR milk data, using rtFMS to reduce human bias. The outcome of this study presents an option for improved prediction models, reducing error due to empirical knowledge or habitual preferences. The implementation of rtFMS to design future prediction models can guide model inputs and features. In future work, improved prediction models can be developed for detecting blood BHBA and blood NEFA in dairy cows using the findings of our research. Our prediction models have the potential to increase early detection of metabolic disorders during the transition period in dairy cows.

Funding

The project was supported by funds of the Federal Ministry of Food and Agriculture (BMEL) based on a decision of the Parliament of the Federal Republic of Germany via the Federal Office for Agriculture and Food (BLE) under the innovation support program.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

The authors acknowledge the MPR Bayern e.V. (Bavarian Association for Raw Milk Testing) and the LKV Bayern e. V. (Dairy Herd Improvement Association of Bavaria) for supporting the collection of our data with their funds. We appreciate the work done by FOSS for providing the fatty acid panels used for the analysis. We gratefully acknowledge all dairy farms taking part in our project and the team of the laboratory in the Clinic for Ruminants, Ludwig- Maximilian-University Munich. Special thanks to our colleagues, Anne Reus, Franziska Hajek and Dr. Stefan Plattner for supporting the data collection. The official number for the animal experiment proposal by the Government of Bavaria was ROB-55.2Vet-2532.Vet_03-17-84 and in the

Thuringian State Office of Consumer Protection was it 22-2684-04-LMU-17-101.

References

Benedet, A., Franzoi, M., Manuelian, C.L., Penasa, M., De Marchi, M., 2020. Variation of blood metabolites of Brown Swiss, Holstein-Friesian, and Simmental cows. Animals 10, 271. https://doi.org/10.3390/ani10020271.

Carrier, J., Stewart, S., Godden, S., Fetrow, J., Rapnicki, P., 2004. Evaluation and use of three cowside tests for detection of subclinical ketosis in early postpartum cows. J. Dairy Sci. 87, 3725–3735. https://doi.org/10.3168/jds.S0022-0302(04)73511-0.

Chapinal, N., LeBlanc, S.J., Carson, M.E., Leslie, K.E., Godden, S., Capel, M., Santos, J.E. P., Overton, M.W., Duffield, T.F., 2012. Herd-level association of serum metabolites in the transition period with disease, milk production, and early lactation reproductive performance. J. Dairy Sci. 95, 5676–5682. https://doi.org/10.3168/jds.2011-5132.

Chollet, F., 2018. Deep Learning with Python. Manning Publications Co, Shelter Island, New York.

Díaz-Pacheco, A., Gonzalez-Bernal, J.A., Reyes-García, C.A., Escalante-Balderas, H.J., 2018. Full model selection in big data. In: Castro, F., Miranda-Jiménez, S., González-Mendoza, M. (Eds.), Advances in Soft Computing, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 279–289. https://doi.org/10.1007/978-3-030-02837-4_23.

Duffield, T.F., Lissemore, K.D., McBride, B.W., Leslie, K.E., 2009. Impact of hyperketonemia in early lactation dairy cows on health and production. J. Dairy Sci. 92, 571–580. https://doi.org/10.3168/jds.2008-1507.

Escalante, H.J., Montes, M., Sucar, L.E., 2009. Particle swarm model selection. J. Mach. Learn. Res. 10, 405–440.

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

Giuliodori, M.J., Delavaud, C., Chilliard, Y., Becú-Villalobos, D., Lacau-Mengido, I., de la Sota, R.L., 2011. High NEFA concentrations around parturition are associated with delayed ovulations in grazing dairy cows. Livest. Sci. 141, 123–128. https://doi.org/10.1016/j.livsci.2011.05.007.

Gordon, J.L., LeBlanc, S.J., Duffield, T.F., 2013. Ketosis treatment in lactating dairy cattle. Vet. Clin. North Am. Food Anim. Pract. 29, 433–445. https://doi.org/10.1016/j.cvfa.2013.03.001. Metabolic Diseases of Dairy Cattle.

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

Gruber, S., Tremblay, M., Kammer, M., Reus, A., Hajek, F., Plattner, S., Baumgartner, C., Hachenberg, S., Döpfer, D., Mansfeld, R., 2021. Evaluation of an early warning system for elevated ß-hydroxy-butyrate and non-esterified fatty acid values based on Fourier transform infrared spectra from routine milk samples. Milk Science International - Milchwissenschaft 74 (2). https://doi.org/10.48435/MSI.2021.2.

Hansen, P.W., 1999. Screening of dairy cows for ketosis by use of infrared spectroscopy and multivariate calibration. J. Dairy Sci. 82, 2005–2010. https://doi.org/10.3168/jds.S0022-0302(99)75437-8.

Heuer, C., Luinge, H.J., Lutz, E.T.G., Schukken, Y.H., van der Maas, J.H., Wilmink, H., Noordhuizen, J.P.T.M., 2001. Determination of acetone in cow milk by fourier transform infrared spectroscopy for the detection of subclinical ketosis. J. Dairy Sci. 84, 575–582. https://doi.org/10.3168/jds.S0022-0302(01)74510-9.

Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. J. Comput. Graph. Stat. 15, 651–674. https://doi.org/10.1198/106186006X133933.

Hothorn, T., Zeileis, A., 2015. Partykit: a modular toolkit for recursive partytioning in R. J. Mach. Learn. Res. 16, 3905–3909.

Iwersen, M., Klein-Jöbstl, D., Pichler, M., Roland, L., Fidlschuster, B., Schwendenwein, I., Drillich, M., 2013. Comparison of 2 electronic cowside tests to detect subclinical ketosis in dairy cows and the influence of the temperature and type of blood sample on the test results. J. Dairy Sci. 96, 7719–7730. https://doi.org/10.3168/jds.2013-7121.

Krstajic, D., Buturovic, L.J., Leahy, D.E., Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. J. Cheminf. 6, 10. https://doi.org/10.1186/1758-2946-6-10.

Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 1 (5), 2008.

Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3.

Mehtiö, T., Mäntysaari, P., Kokkonen, T., Kajava, S., Latomäki, T., Nyholm, L., Grelet, C., Pitkänen, T., Mäntysaari, E., Lidauer, M., 2018. Developing an indicator for body fat mobilisation using mid-infrared spectrometry of milk samples in dairy cows. In: Proc. World Congr. Genet. Appl. Livest. Prod. Electronic Poster Session-Biology-Feed Intake and Efficiency, 1, p. 225.

Oetzel, G.R., 2004. Monitoring and testing dairy herds for metabolic disease. Vet. Clin. North Am. Food Anim. Pract. 20, 651–674. https://doi.org/10.1016/j.cvfa.2004.06.006.

Paradkar, M.M., Irudayaraj, J., 2002. Determination of cholesterol in dairy products using infrared techniques: 1. FTIR spectroscopy. Int. J. Dairy Technol. 55, 127–132. https://doi.org/10.1046/j.1471-0307.2002.00044.x.

Pralle, R.S., Weigel, K.W., White, H.M., 2018. Predicting blood β-hydroxybutyrate using milk Fourier transform infrared spectrum, milk composition, and producer-reported variables with multiple linear regression, partial least squares regression, and artificial neural network. J. Dairy Sci. 101, 4378–4387. https://doi.org/10.3168/jds.2017-14076.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rieppo, L., Saarakkala, S., Närhi, T., Helminen, H.J., Jurvelin, J.S., Rieppo, J., 2012. Application of second derivative spectroscopy for increasing molecular specificity of fourier transform infrared spectroscopic imaging of articular cartilage. Osteoarthr. Cartil. 20, 451–459. https://doi.org/10.1016/j.joca.2012.01.010.

Rohman, A., Setyaningrum, D.L., Riyanto, S., 2014. FTIR spectroscopy combined with partial least square for analysis of red fruit oil in ternary mixture system. Int. J. Spectrosc. https://doi.org/10.1155/2014/785914.

Seifi, H.A., LeBlanc, S.J., Leslie, K.E., Duffield, T.F., 2011. Metabolic predictors of post-partum disease and culling risk in dairy cattle. Vet. J. 188, 216–220. https://doi.org/10.1016/j.tvjl.2010.04.007.

Sim, S.F., Jeffrey Kimura, A.L., 2019. Partial Least Squares (PLS) Integrated Fourier Transform Infrared (FTIR) approach for prediction of moisture in transformer oil and lubricating oil. J. Spectrosc. https://doi.org/10.1155/2019/5916506.

Tiplady, K.M., Sherlock, R.G., Littlejohn, M.D., Pryce, J.E., Davis, S.R., Garrick, D.J., Spelman, R.J., Harris, B.L., 2019. Strategies for noise reduction and standardization of milk mid-infrared spectra from dairy cattle. J. Dairy Sci. 102, 6357–6372. https://doi.org/10.3168/jds.2018-16144.

Tremblay, M., Kammer, M., Lange, H., Plattner, S., Baumgartner, C., Stegeman, J., Duda, J., Mansfeld, R., Döpfer, D., 2018. Identifying poor metabolic adaptation during early lactation in dairy cows using cluster analysis. J. Dairy Sci. 101, 7311–7321.

Tremblay, M., Kammer, M., Lange, H., Plattner, S., Baumgartner, C., Stegeman, J.A., Duda, J., Mansfeld, R., Döpfer, D., 2019. Prediction model optimization using full model selection with regression trees demonstrated with FTIR data from bovine milk. Prev. Vet. Med. 163, 14–23. https://doi.org/10.1016/j.prevetmed.2018.12.012.

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