J. Dairy Sci. 107:10724–10737
https://doi.org/10.3168/jds.2024-25090

Suitability of different machine learning algorithms for the classification of the proportion of grassland-based forages at the herd level using mid-infrared spectral information from routine milk control

A. Birkinshaw,[1] M. Sutter,[1] M. Nussbaum,[1,2] M. Kreuzer,[3] and B. Reidy[1]*
[1]Bern University of Applied Sciences (BFH), School for Agricultural, Forest and Food Sciences (HAFL), CH-3052 Zollikofen, Switzerland
[2]University of Utrecht, Department of Physical Geography, 3584 CB Utrecht, the Netherlands
[3]ETH Zurich, Institute of Agricultural Sciences, CH-8315 Lindau, Switzerland

ABSTRACT

As the call for an international standard for milk from grassland-based production systems continues to grow, so too do the monitoring and evaluation policies surrounding this topic. Individual stipulations by countries and milk producers to market their milk under their own grass-fed labels include a compulsory number of grazing days per year (ranging from 120 d for certain labels to 180 d for others), a specified amount of herbage in the diet, or a prescribed dietary proportion of grassland-based forages (GBF) fed and produced on-farm. As these multifarious policy and label requirements are laborious and costly to monitor on-farm, fast economical proxies would be advantageous to verify the proportion of GBF consumed by the cows in the final product. With this in mind, we employed readily available mid-infrared spectral data (n = 1,132 spectra) from routine milk controls to develop binary classification models for 4 main feed groups from a primarily forage-based diet: total GBF (≥50% [n = 955], ≥75% [n = 599], ≥85% [n = 356]), pasture (≥20% [n = 451], ≥50% [n = 284], ≥70% [n = 152]), fresh herbage (pasture + fresh herbage indoor feeding; ≥20% [n = 517], ≥50% [n = 325], ≥70% [n = 182]), and whole plant corn (fresh + conserved; ≥10% [n = 646], ≥30% [n = 187]), with the latter as a negative control. We compared 4 machine learning methods to assess which statistical model performs best at discriminating these classes. Three of these models have not yet been tested for herd-level dietary proportion classification, and all 4 follow completely different approaches: least absolute shrinkage and selection operator (LASSO), partial least squares discriminant analysis (PLS-DA), random forest (RF), and support vector machines (SVM). Seasonality has been a missing element from previous dietary herbage proportion classification models. As grazing and fresh herbage indoor feeding are highly dependent on the season, we developed an indicator to incorporate seasonality in a consistent, unbiased manner into our models. We also tested 3 sets of covariates. The first set included only mid-infrared spectra derived data, the second included mid-infrared spectra derived data plus seasonality indices and the third included mid-infrared spectra derived data, seasonality indices and additional herd specific information (DIM, breed, and parity). Of the 4 machine learning algorithms tested for the binary classification of GBF proportion at herd level, LASSO and PLS-DA performed best according to evaluation metrics; however, the RF and SVM models were not far behind the best performing model evaluation metrics in each feed category. Our best performing model, the LASSO model containing seasonality indices and herd specific information, classified total GBF ≥50% with an accuracy of 78.6%, precision of 85.1%, sensitivity of 90.6%, specificity of 14.1%, and F1 score (harmonic mean of precision and sensitivity) of 87.7%; this was very similar to the PLS-DA model. Our results suggest that in general, LASSO and PLS-DA machine learning algorithms perform better for dietary GBF classification than RF or SVM algorithms.

Key words: machine learning, mid-infrared spectroscopy, grassland-based feeding, seasonality index

INTRODUCTION

In recent years, the application of machine learning algorithms has become increasingly popular in dairy research (Shine and Murphy, 2022). This largely emanates from the demand for rapid analyses of dairy products to determine farming practices, coming from the dairy industry supplying the growing sector of conscious consumers. One of the main drivers is the call to establish an

Received April 24, 2024.
Accepted July 26, 2024.
*Corresponding author. beat.reidy@bfh.ch

The list of standard abbreviations for JDS is available at adsa.org/jds-abbreviations-24. Nonstandard abbreviations are available in the Notes.

internationally recognized grass-based dairy cow feeding standard (Alothman et al., 2019; Moscovici Joubran et al., 2021; Birkinshaw et al., 2023). The use of grassland-based forages (**GBF**), including fresh and conserved herbage as silage or hay, is often encouraged in countries with temperate climates and abundant herbage growth (Moscovici Joubran et al., 2021; FOAG, 2022). This is because grasslands have proven to be climate neutral despite increasing livestock numbers and management intensification (Chang et al., 2021). Conscious consumers perceive GBF to be superior to conventional milk production systems in terms of increased animal welfare and the nutritional value of the milk (Infascelli et al., 2021), and they are often willing to pay a premium for "grass-fed" products (Stampa et al., 2020). The milk fat from GBF production systems is richer in beneficial fatty acids such as n-3 and vaccenic acid, whereas less desirable fatty acids, such as n-6 and palmitic acid, are reduced (Alothman et al., 2019, Moscovici Joubran et al., 2021), allowing GBF production to be differentiated from other production systems.

Progressively more countries are imposing compulsory annual grazing periods, with milk producers following suit to market their milk under their own grass-fed labels. In Switzerland, for example, the Regular Outdoor Exercise Program endorsed by the Federal Office of Agriculture (**FOAG**) requires ruminants be on pasture for 26 d of the month from May to October (FOAG, 2023). Milk labels such as "Weidemelk" in the Netherlands, "Organic Valley" in the United States, and "Marguerite Happy Cow" in Belgium require at least 120, 150, and 180 d of grazing per year, respectively (Weidemelk, 2022; Hill, 2023; Marguerite Happy Cow, 2015). Of course, allowing pasture access to cows does not guarantee substantial herbage intake, as grazing behavior is affected by multiple factors including pasture management (Pérez-Prieto and Delagarde, 2012), climatic conditions, diurnal cycles (Legrand et al., 2009), and even individual cow preferences (preferred plants and familiar housing/surroundings, Rutter, 2010). In an effort to avoid these sources of variation, GBF initiatives are often multifactorial. As an example, Bord Bia, the Irish Food Board, has recently established a grass-fed dairy standard stipulating that at least 90% of a cow's diet be composed of herbage or herbage-based forage (on a fresh weight basis), with at least 160 d on pasture (deducting 80 d from the national grazing day average of 240 d; Bord Bia, 2019). These multifarious policy and label requirements are undoubtedly laborious to monitor on-farm. Instead, fast and economically viable proxies are required to verify the proportion of GBF consumed using milk quality traits; this is easiest by evaluating the end product (milk) for dietary biomarkers.

Mid-infrared (**MIR**) spectroscopy has proven a strong competitor among dairy measurement approximation approaches as it is rapid, cost-effective, and nondestructive (Pereira et al., 2020). Milk payment systems make use of this technology globally for routine assessment of contents of fat, protein, lactose and urea in the milk (Tiplady et al., 2020). Over the last decade, milk MIR spectroscopy has been used not only to differentiate distinct dietary classes (Frizzarin et al., 2021 and 2023) and management systems (Capuano et al., 2014; Soyeurt et al., 2022), but also to classify the proportions of specific feed components in the ration (Coppa et al., 2021). Capuano et al. (2014) were the first to show that MIR spectroscopy could be used to differentiate feeding regimens (grass vs. no grass), and more recently, Soyeurt et al. (2022) introduced the seasonality of GBF into their classification models by including the sample month as an indirect indicator of GBF intake. May to August were considered full grazing periods, whereas March, April, September, and October were considered transitionary periods, and the winter months (November, December, January, February) were classified as periods that were not grass-based (Soyeurt et al., 2022). Capuano et al. (2014) and Coppa et al. (2021) used MIR spectra for their binary classification models, whereas Soyeurt et al. (2022) used 4 predictors given directly by the spectrometer and 44 predictors derived from previously developed equations rather than the spectral wavenumbers themselves.

As previous studies aiming at the prediction of binary proportions of GBF and other dietary components have been limited to partial least squares discriminant analysis (**PLS-DA**), it remains unclear which statistical model performs best in this context and whether the choice of model is critical. Moreover, the study that included binary proportions of GBF and other feed components in their models (Coppa et al., 2021) did not include predicted milk quality traits from MIR spectra, only the wavenumbers themselves. Of the studies that included predicted milk quality traits (Capuano et al., 2014; Soyeurt et al., 2022), one included only predicted milk quality traits (no wavenumbers); both predicted grass presence or absence, not the binary GBF proportions themselves; and one did this without any dietary information, but by basing herbage intake on an indirect indicator, the month of the year (Soyeurt et al., 2022). To our best knowledge, no previous studies have attempted to develop and test a reproducible seasonality index that was included in their models. A multifaceted approach, combing these aspects, is pertinent in the development of a reliable, internationally recognized grassland-based feeding standard that can be applied on at the herd level using readily available bulk tank milk MIR spectral data.

Therefore, the present study had 2 aims. The first aim was to compare the predictive performance of 4 different machine learning algorithms (regularized generalized linear [least absolute shrinkage and selection operator; **LASSO**] models, PLS-DA, random forest [**RF**], and support vector machines [**SVM**]) by using a multifaceted prediction approach including bulk tank milk MIR predictors given directly by the spectrometer (protein and fat) and selected MIR wavenumbers to classify specified proportions of 3 different categories of GBF. These included total GBF (all herbage, fresh + conserved), pasture (grazed herbage), fresh herbage (pasture + fresh herbage indoor feeding), and whole plant corn (fresh + conserved) as a negative control. The second aim was to develop and test a reproducible seasonality index that can be included in future GBF predictions.

MATERIALS AND METHODS

Farm, Animal, and Feed Ration Data

The present study was conducted in Switzerland from January 2020 to December 2021. The duration of 2 calendar years was intentionally chosen to reflect and incorporate seasonal variations in feeding strategies. We selected 27 dairy farms as a convenience sample to represent a wide range of feeding and farming strategies based on (1) contrasting proportions of GBF (24%–100%), fresh herbage (pasture and fresh herbage, 0%–100%), and whole plant maize (fresh and conserved, 0%–52%) in the diet; (2) different dairy breeds; (3) calving strategies (seasonal or not); and (4) diverse locations within Switzerland (Supplemental Figure S1, see Notes). Each MIR spectrum was classified according to the herd's diet composition at a certain time point on the respective farms not according to the farm. We selected 27 farms to give a wide range of diet compositions that changed on a monthly basis.

All farms signed informed consent for inclusion in the study. A priori inclusion criteria required farms be registered with a breeding association as this ensured access to bulk tank milk quantity and quality data at 11 equally distributed points in time throughout the year. Lactating herd size, parity, DIM, and dairy breed information for each month were accessed via the respective breeding association and the national animal tracing database (**TVD**; FOAG, 2019; Table 1). A dominant dairy breed was assigned to each farm according to the majority (>80%) of the lactating herd. Accordingly, the breeds on 12 farms were classified as Brown Swiss + Jersey (Jersey was combined with Brown Swiss because only 1 farm kept Jersey cows), 12 were classified as Simmental (including Simmental × Red Holstein), and 3 were classified as Holstein.

Table 1. Dairy herd characteristics, performance, and milk quality traits (n = 27 farms)

Item	Mean	Median	Minimum	Maximum	SD
Lactating herd (head)[1]	32.4	26.8	6.01	73.7	16.2
Parity (n)	3.09	2.98	1.74	6.00	0.70
DIM	165	169	33.1	290	48.2
Milk yield (kg/d)	23.3	22.3	12.0	35.7	4.82
ECM (kg/d)	24.1	23.0	13.2	36.5	4.98
Milk constituents					
Fat (%)	4.26	4.23	3.50	5.77	0.302
Protein (%)	3.46	3.43	2.85	4.38	0.205
SCC (10^3/mL)	134	106	8	1,001	92

[1]Calculated monthly (based on the calving date and dry period) from daily herd structure data thus considering the exact dates when cows were leaving or joining the herd.

Dietary proportions were obtained and verified using a series of cross-checks. In the first step, average DM intake was calculated on herd level according to the average ECM production, BW, DIM, and parity (Jans et al., 2015). The average calculated amounts of dietary DM to be consumed were compared with farmer reported rations surveyed during a telephone interview and an on-farm visit during the winters of 2020 and 2021. In the next step, purchase receipts and manufacturer labels were used to determine the amount (in kg) and nutritional information of concentrates and other feeds (such as sugar beet pulp and brewers grains) purchased off-farm. Forage (hay, lucerne, grass silage, and whole plant corn [fresh and conserved]) produced on-farm was expressed in quantities per year. Farmers reported monthly intakes of grass silage, hay, and whole plant corn (fresh and conserved) in percentages of the total ration, and the remainder was attributed to fresh herbage intake (grazed pasture and fresh herbage fed indoors). The reported monthly fed quantities were then compared with the reported monthly supplied quantities (purchased and produced on-farm) by means of a balance sheet. The Swiss Feed Database (Agroscope, 2016) was employed to calculate contents of NE_L, absorbable protein in the duodenum (based on energy), and absorbable protein in the duodenum (based on nitrogen) of the feed components produced and fed on-farm. Lastly, total amounts of NE_L and absorbable protein in the duodenum based on energy and nitrogen were calculated for the reported ration and cross-checked with the average requirements for the lactating herd according to Jans et al. (2015). A discrepancy of ≤10% was considered plausible.

Milk MIR Spectroscopy

In Switzerland, bulk tank milk samples are collected twice per month (on random days) for routine quality control and milk payments by the national milk testing laboratory (Suisselab AG, Zollikofen, Switzerland).

Milk fat and protein are analyzed with MIR spectroscopy (MilkoScan FT6000, Foss, Hillerød, Denmark), and the spectral data can be stored. Tested bulk tank milk samples consisted of evening and subsequent morning milk stored at <4°C until processing within the next 24 h. The Foss MIR spectra contain 1,060 infrared frequencies that represent the absorption of infrared light through the raw milk sample at wavelengths in the 925 to 5,008/cm range. Pin numbers obtained directly from the MilkoScan FT6000 were converted to wavenumbers by multiplying by 3.858 (Dini et al., 2021), and only the spectral regions relevant to feed intake were included in the models (2,989−2,561/cm; 1,809−1,712/cm; 1,600−926/cm) according to Maurice-Van Eijndhoven et al. (2013) and Coppa et al. (2021), resulting in 1,199 wavenumbers and 55 actual data points at a spectral resolution of 22/cm). To provide reproducible, internationally comparable models for GBF classification, all MIR spectra were standardized according to the OptimMIR standardization protocol (Grelet et al., 2015). From an initial 1,296 spectra, 164 (12.7%) were excluded because these cases lacked some data or they were considered outliers (based on a Z-score of >3 or <−3). For analysis, 1,132 MIR spectra were included, 578 from 2020 and 554 from 2021. These cases included all milk, farm, animal, and diet data.

Seasonality and Additional Covariates

The date on which each milk sample was collected contained information about seasonal variations of the main feed components in the herds' diets and the stage of lactation. Because we were interested in the reoccurring seasonality pattern instead of the absolute date, we created seasonality indices by transforming the number of days in the year trigonometrically (as often done with terrain aspect in environmental mapping, e.g., Amatulli et al., 2018). Each sampling date was plotted on a circle representing the entire sampling year (Figure 1), starting with January 1 at the top and ending with December 31 in the same position. The position of each sampling date was then transferred to an angle α in radians to compute 2 continuous indices:

$$\text{Brumal index: } I_b = cos(\alpha), \text{ and}$$

$$\text{Vernal index: } I_v = sin(\alpha).$$

The Brumal index (I_b) yields values of 1 for January 1 and −1 for June 31 and results in values near 0 for spring and autumn. The Vernal index (I_v) results in values of 1 for March 31 and −1 for September 30. Creating seasonality indices by the transformation of sample dates in this manner ensured that day, month, and season were included systematically in model training by forming continuous variables bridging the value offset between December and January; it also allows for consistent, unbiased prediction of samples taken at future dates.

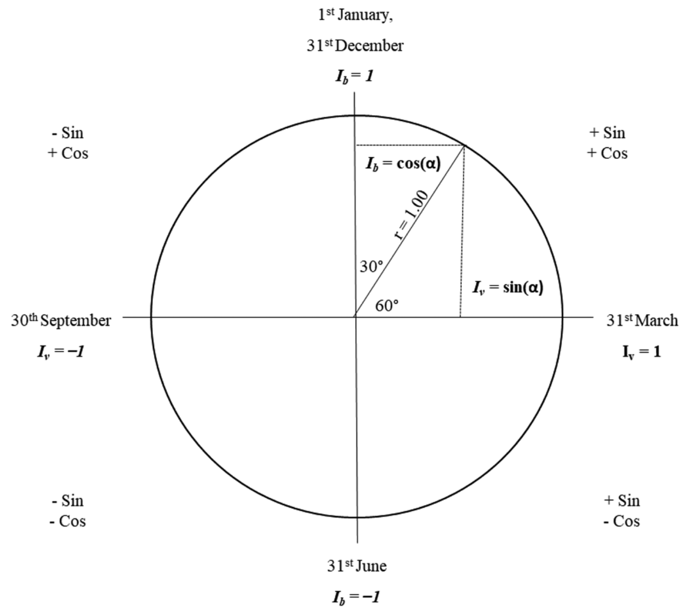

Figure 1. Seasonality indices created by transforming the milk sample dates trigonometrically to 2 continuous indices: Brumal (I_b) and Vernal (I_v).

In addition to the milk MIR spectra, seasonality indices, and quality traits (milk fat and protein contents), we considered additional explanatory factors. Breed, parity, and DIM had been obtained for each herd, every month, from the respective breeding association and the TVD. We chose to include the wavenumbers themselves as well as the predicted fat and protein derived from MIR spectra even though they may be correlated (as done in feature engineering), as more information directly relating to the response rather than the original data may be present.

Model Development

We developed binary classification models based on a range of plausible grassland-based feeding standards (total GBF: ≥50%, ≥75%, ≥85%; pasture: ≥20%, ≥50%, ≥70%; fresh herbage [pasture and indoors]: ≥20%, ≥50%, ≥70%; and whole plant corn [fresh and conserved]: ≥10%, ≥30%) using 3 sets of explanatory covariates: (1) contained only MIR spectra and milk quality traits (milk fat and protein); (2) contained MIR spectra and milk quality traits with the addition of the seasonality indices (Brumal and Vernal indices derived from the sample date); and (3) contained MIR spectra, milk quality traits, the seasonality indices, and breed, parity, and DIM as additional covariates. The classification of total GBF ≥75% or ≥85% corresponds to the Grassland-Based Milk and Meat Production program implemented by FOAG (2024), whereas the pasture (and fresh herbage) classifications ≥20% and ≥70% relate to another 2 of the animal welfare feeding programs encouraged by FOAG (2023), the Regular Outdoor Exercise Program and the "Weidebeitrag" (grazing contribution), respectively. The whole plant corn classifications we chose were loosely based on the remainder of the dietary intake once certain FOAG requirements have been fulfilled.

It should be noted that in contrast to our other covariates (MIR spectra, milk quality traits, and the seasonality indices), the data for the additional covariates originates from the corresponding Swiss breeding associations; thus, they are not as readily available as the other covariates. Therefore, we wanted to test the classification capability of our models by including and excluding this information to determine whether the additional data collection effort is beneficial. As dietary data from each farm was recorded once per month and milk samples were collected twice, 2 sets of MIR spectra from the same month were linked to only 1 set of dietary information.

We used the Boruta algorithm (Boruta package, version 8.0.0; Kursa and Rudnicki, 2010) for initial determination of the most important wavenumbers (from the spectral regions relevant to feed intake; Maurice-Van Eijndhoven et al., 2013; Coppa et al., 2021) for each of the 4 feed groups. Boruta is a feature selection method that maintains only covariates (in this case, wavenumbers) that are important (Kursa, 2022). This is done by duplicating the wavenumber data and shuffling each wavenumber randomly (called shadow features or wavenumbers in our case). The randomly permuted wavenumbers are appended to the original data and a RF model (see the "Prediction Methods" section) is fitted. Wavenumber importance is evaluated by Mean Decrease Accuracy and only the wavenumbers that exhibit larger importance than the average importance of shadow wavenumbers are retained. This resulted in a unique set of important wavenumbers for each feed group that was used in all subsequent models.

Validation of Model Performance

Farm-based cross-validation was applied for all 4 approaches and all 3 sets of covariates. To ensure an unbiased estimation of validation statistics, data from 3 randomly chosen farms (10% of the data set) were set aside at each iteration as test sets, and data from the remaining 24 farms were included in the training set. The following confusion matrix metrics were used to evaluate the classification performance of each model, where n = total cases:

$$\text{accuracy} = \frac{\text{true negative} + \text{true positive}}{\text{n}}, \quad [1]$$

$$\text{precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}, \quad [2]$$

$$\text{sensitivity/recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}, \quad [3]$$

$$\text{specificity} = \frac{\text{true negative}}{\text{true negative} + \text{false positive}}, \quad [4]$$

$$\text{F1 score} = \frac{2 \times \text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}}. \quad [5]$$

Prediction Methods

We chose to compare 4 methods that follow different approaches to establish response-covariate relationships. The first method, LASSO, is based on the generalized linear model framework; the second method, PLS-DA, is a multivariate dimensionality reduction tool that was applied in the only other similar study we are aware of (Coppa et al., 2021). The third method, RF, is based on stepwise binary splitting of the data, whereas the fourth method, SVM, applies a high dimensional kernel function to the data.

LASSO. Least absolute shrinkage and selection operator is an extension of the popular generalized linear models with an additional constraint in the coefficient estimation. Through its shrinkage method, LASSO has a built-in variable selection that helps to deal with small sample sizes and correlated data, making it an attractive framework for high-dimensional data. The LASSO method estimates coefficients of a linear model by minimizing a penalized residual sum of squares, with the penalty being equal to the weighted sum of absolute values of coefficients. The optimal weight, i.e., degree of so-called regularization is found by computing the regularization path with increasing the penalty along increasing values for λ, the regularization parameter (Hastie and Qian, 2006). Therefore, higher λ values lead to stronger regularization, and, as such, a kind of continuous subset selection is performed. Coefficients shrunken to zero are excluded from the model (Hastie et al., 2009). The optimal λ was found for our models based on a sequence of 10 λ values along a range with no penalty applied, to a penalty for which all coefficients are zero. Variable importance was computed via permutation.

PLS-DA. Partial least squares discriminant analysis is derived from PLS regression. In binary classification models, this linear, 2-class classifier aims to find a straight line that divides the space into 2 regions by finding a discriminator, separator, or decision function (Brereton and Lloyd, 2014). In theory, PLS-DA combines dimensionality reduction and discriminant analysis into a single algorithm that handles high dimensional and highly correlated data particularly well (Lee et al., 2018). In the present study, we employed the PLS1-DA (Lee et al., 2018) algorithm for our binary classification models. We set the number of components for the model to be trained on from 2 to 5. This allowed the system to automatically evaluate a range of candidate models. Accuracy was used to select the optimal model using the largest value.

RF. Random forest grows a large number of decorrelated decision or regression trees and combines the predictions by taking the average (Hastie et al., 2001). The basic principle underlying RF is that averaging predictions from many uncorrelated models outperforms 1 single model. Each tree is independently fitted to a data set randomly sampled (with replacement) from the training data resulting in the same response distribution for all trees in the forest (Williams et al., 2020). Trees are de-correlated by randomly selecting only a subset of variables (tuning parameter m_{try}) to test as split criteria at each node of the decision or regression trees. Random forest performs equally well on small sample sizes and high dimensional data (Biau and Scornet, 2016). The R package "Ranger" (Wright and Ziegler, 2017) offers a fast implementation of RF (Breiman, 2001). We used default values for the number of decision trees (n = 500) and m_{try}, the square root of the total number of covariates. Covariate importance was determined by permutation, which is computed by evaluating the increase in prediction error when values of a variable were randomly shuffled (Molnar, 2022).

SVM. The SVM method originates from binary classification problems and divides a high dimensional space spanned by covariates into 2 sections according to the response class observations (James et al., 2013). Some of the observations are allowed to be on the wrong side of the margins of this division (margin of tolerance C). The division may be done by linear functions resulting in hyperplanes or more complex nonlinear functions such as Gaussian kernels. We used radial basis functions, and the tuning parameters C and Σ (degree of kernel smoothness) were optimized in a 2-step procedure. First, we applied a broad range of 9 different values for Σ and C. In a second step, we used a fine grain grid around the found parameters of the first step to ultimately determine Σ and C.

For LASSO, PLS-DA, and SVM, data were centered and scaled before model fitting. Random forest does not require such preprocessing, as each covariate is separately fitted to the data at each tree split, and different data ranges do not distort models with numerous covariates.

Calculations and Software

The ECM amount (in kg; Table 1) was calculated according to Jans et al. (2015):

$$\mathrm{ECM} = \mathrm{milk}\left(\mathrm{kg}\right)\times$$
$$\left[0.38\times\mathrm{fat}\left(\%\right)+0.24\times\mathrm{protein}\left(\%\right)+0.17\times\mathrm{lactose}\left(\%\right)\right]\Big/3.14.$$

Protein values were given as true protein, and lactose was expressed as lactose monohydrate in Foss Milkoscan predictions. A standard value of 4.8% lactose was applied in the above equation. All statistical analyses were performed with the statistical software R (version 4.1.0; R Core Team, 2022). Model training and cross-validation were aided by the caret package (version 6.0.93; Kuhn, 2008) for RF (ranger, version 0.14.1; Wright and Zeigler, 2017), SVM (svmRadial, kernlab, version 0.9.31; Karatzoglou et al., 2022), LASSO (glmnet, LASSO, version 4.1.4; Freidman et al., 2010), and PLS-DA (pls, version 2.8-2; Liland et al., 2023) machine learning algorithms. Our seasonality indices were created by transforming the milk sample date into 2 continuous indices, Brumal and Vernal, with the circular package (version 0.4-95, Agostinelli and Lund, 2022). Test sets were created by randomizing data into 9 test groups (of 3 farms each) using the greedy method from groupdata2 (version 2.0.1; Olsen, 2022). Maps were created with bfsMaps (version 0.9.8; Signorell, 2022) and plots with ggplot2 (version 3.4.0; Wickham, 2016).

RESULTS

Farm Characteristics

Average dairy herd characteristics (lactating herd, parity, and DIM), performance, and milk quality traits are presented in Table 1. The overall mean milk fat and protein contents of the milk samples included in our classification models were 4.26% and 3.46%, respectively. Milk yield (mean = 23.3 kg/d) and ECM yield (mean = 24.1 kg/d) are included in Table 1 as additional information. Table 2 contains reported and verified information on herd intake per head and diet composition. Mean DMI was 19.0 kg/d per cow, and diets were primarily forage-based with an average of 9.6% and 3% of the diet derived from concentrates and other feeds, respectively.

Table 2. Dairy herd intake and diet composition (n = 27 farms)

Item	Mean	Median	Minimum	Maximum	SD
DMI (kg/head per d)	19.0	18.7	13.9	23.3	1.9
Feed component (%)					
Total grassland-based forage	73.9	76.6	24.2	100	19.1
Pasture	25.3	0	0	100	32.4
Fresh herbage (pasture + fresh herbage indoor feeding)	29.1	11.5	0	100	33.8
Whole plant corn	13.6	10.8	0	52.1	13.6
Other feeds	3.0	0	0	39.5	6.8
Concentrates	9.6	8.5	0	34.9	7.3

Performance of Classification Models

Results of the 4 different classification methods (LASSO, PLS, RF, and SVM) for each forage are displayed in Table 3. The accuracy, precision, sensitivity (recall), specificity, and F1 score are presented for all models. Detailed herd characteristics, performance, milk quality traits, and DMI and dietary proportions, specified for each of the dietary groups, can be found in Supplemental Tables S1, S2, S3, and S4 (see Notes).

Total Grassland-Based Forages

The LASSO models provided the highest accuracy for total GBF ≥75% and ≥85% (Table 3). For total GBF ≥50%, the PLS-DA model was the most accurate. The RF models were the least accurate when classifying total GBF ≥75% and ≥85%, whereas the LASSO model was the least accurate for total GBF ≥50%. In terms of precision, the LASSO models classified total GBF ≥75% and ≥85% with the highest precision of the 4 models, with no difference among the 4 models for total GBF ≥50% (Table 3). The RF models were the least precise when classifying total GBF ≥75% and ≥85%. Sensitivity was highest in the LASSO models classifying total GBF ≥75% and ≥85%, and in the PLS-DA model classifying total GBF ≥50%. When classifying total GBF ≥50%, the LASSO model was the least sensitive; for total GBF ≥75% and ≥85%, the SVM and RF models were the least sensitive, respectively (Table 3). Specificity was generally very low for all models classifying total GBF ≥50%, with the PLS-DA model performing the worst. The LASSO and PLS-DA models demonstrated the highest specificity for classifying total GBF ≥75% and ≥85%, respectively, whereas the RF and SVM models were the least specific for classifying ≥75% and ≥85% total GBF proportions, respectively. The F1 scores were highest in the PLS-DA model classifying total GBF ≥50% and the LASSO models classifying total GBF ≥75% and ≥85%; they were lowest in the LASSO model classifying total GBF ≥50% and the RF/SVM and RF models classifying total GBF ≥75% and ≥85% (Table 3). The addition of seasonality indices did not greatly improve the classification capabilities of total GBF across models, and the best improvement was by 2 percentage points in the specificity of the SVM model classifying total GBF ≥75%. Generally, the same trend was observed when adding the breeding association data (DIM, breed and parity); however, an improvement of 7% was recorded in the very low specificity of the LASSO model classifying total GBF ≥50% (Table 3). The 20 most important variables for the best performing models, in terms of precision (LASSO model classifying total GBF ≥50% with all 3 sets of covariates) and sensitivity (PLS-DA model classifying total GBF ≥50% with seasonality indices) are displayed in Figures 2 and 3, respectively. Wavenumbers in the C–H bending and C–O stretching regions (900–1,500/cm) of the MIR spectra make up the majority of the 20 most important variables in total GBF classification (Figures 2 and 3, Total GBF). The classification of total GBF ≥50% was better discriminated than the classification of the 2 other GBF categories (≥75% and ≥85%; Table 3).

Pasture

The LASSO models were the most accurate for classifying pasture ≥20%, ≥50%, and ≥70%, whereas the RF models were the least accurate for classifying all 3 categories of dietary pasture intake. Precision was highest in the LASSO models when classifying pasture intake ≥20% and ≥50%, and in the PLS-DA model when classifying pasture ≥70%. Precision was lowest for all 3 categories of pasture intake classification with the RF models. In terms of sensitivity, the LASSO models performed best across all 3 pasture intake classification categories, and the RF models performed the worst. The PLS-DA models registered the highest specificity across all 3 pasture classification categories, and the RF models registered the lowest for pasture classification intake ≥20% and ≥50%. For classification of pasture ≥70%, the SVM model performed the poorest. The F1 scores were highest for the LASSO models for all 3 pasture classification categories and lowest for the RF models across the 3 pasture classification categories. Generally, the addition of the seasonality indices to the models made little difference; however, in a couple of models, the differences were

Table 3. Validation statistics for 4 different classification algorithms and 3 sets of covariates (MIR data, MIR data + seasonality indices, and MIR data + seasonality indices + DIM + breed + parity) for herd-level grassland-based feed and whole plant corn composition (negative control) computed by farm-based cross-validation (n = 1,132 samples/spectra)[1]

Feed component (%)	Algorithm	MIR data					MIR data + seasonality indices					MIR data + seasonality indices + DIM + breed + parity				
		Accuracy	Precision	Sensitivity[2]	Specificity	F1	Accuracy	Precision	Sensitivity[2]	Specificity	F1	Accuracy	Precision	Sensitivity[2]	Specificity	F1
Total GBF																
≥50, n = 955	LASSO	0.79	0.84	0.92	0.07	0.88	0.79	0.84	0.92	0.07	0.88	0.79	0.85	0.91	0.14	0.88
	PLS-DA	0.84	0.84	0.99	0.01	0.91	0.84	0.84	0.99	0.06	0.91	0.83	0.84	0.98	0.02	0.91
	RF	0.82	0.84	0.97	0.02	0.90	0.82	0.84	0.97	0.01	0.90	0.83	0.84	0.98	0.01	0.90
	SVM	0.81	0.84	0.95	0.04	0.89	0.81	0.84	0.95	0.04	0.89	0.80	0.84	0.95	0.00	0.89
≥75, n = 599	LASSO	0.66	0.67	0.69	0.63	0.68	0.66	0.67	0.71	0.62	0.69	0.63	0.65	0.65	0.60	0.65
	PLS-DA	0.60	0.61	0.65	0.53	0.63	0.58	0.59	0.63	0.51	0.61	0.60	0.61	0.68	0.51	0.64
	RF	0.56	0.57	0.66	0.44	0.61	0.56	0.56	0.66	0.45	0.62	0.57	0.58	0.69	0.44	0.63
	SVM	0.59	0.61	0.62	0.55	0.61	0.59	0.62	0.62	0.57	0.62	0.58	0.61	0.60	0.56	0.60
≥85, n = 356	LASSO	0.73	0.58	0.51	0.83	0.55	0.73	0.59	0.51	0.84	0.54	0.68	0.48	0.37	0.81	0.42
	PLS-DA	0.68	0.48	0.26	0.87	0.34	0.67	0.46	0.25	0.86	0.33	0.66	0.44	0.28	0.84	0.34
	RF	0.64	0.37	0.19	0.85	0.26	0.64	0.36	0.18	0.85	0.24	0.66	0.42	0.22	0.87	0.29
	SVM	0.65	0.44	0.41	0.76	0.42	0.65	0.43	0.37	0.77	0.40	0.64	0.42	0.35	0.78	0.38
Pasture																
≥20, n = 451	LASSO	0.75	0.69	0.70	0.79	0.69	0.75	0.68	0.70	0.78	0.69	0.71	0.64	0.65	0.76	0.64
	PLS-DA	0.72	0.67	0.58	0.81	0.62	0.72	0.67	0.58	0.81	0.62	0.70	0.64	0.58	0.78	0.61
	RF	0.65	0.57	0.52	0.73	0.54	0.66	0.58	0.53	0.74	0.55	0.67	0.59	0.55	0.75	0.57
	SVM	0.71	0.63	0.62	0.76	0.63	0.69	0.62	0.61	0.75	0.62	0.66	0.56	0.59	0.70	0.58
≥50, n = 284	LASSO	0.81	0.65	0.53	0.90	0.58	0.81	0.64	0.52	0.90	0.58	0.80	0.61	0.59	0.87	0.60
	PLS-DA	0.79	0.65	0.38	0.93	0.48	0.80	0.68	0.39	0.94	0.49	0.77	0.56	0.43	0.89	0.49
	RF	0.73	0.44	0.25	0.89	0.32	0.73	0.43	0.24	0.89	0.31	0.73	0.42	0.25	0.88	0.31
	SVM	0.78	0.58	0.39	0.91	0.47	0.77	0.57	0.39	0.90	0.46	0.75	0.51	0.54	0.83	0.52
≥70, n = 152	LASSO	0.88	0.63	0.28	0.97	0.39	0.87	0.55	0.18	0.98	0.28	0.87	0.50	0.30	0.95	0.37
	PLS-DA	0.87	0.75	0.04	0.99	0.08	0.87	0.80	0.03	0.99	0.05	0.86	0.17	0.01	0.99	0.02
	RF	0.84	0.14	0.03	0.97	0.05	0.84	0.07	0.01	0.97	0.02	0.84	0.09	0.02	0.97	0.03
	SVM	0.85	0.33	0.13	0.96	0.19	0.85	0.36	0.11	0.97	0.17	0.84	0.33	0.22	0.93	0.26
Fresh herbage																
≥20, n = 517	LASSO	0.78	0.75	0.76	0.79	0.76	0.77	0.75	0.75	0.79	0.75	0.76	0.74	0.72	0.79	0.73
	PLS-DA	0.75	0.73	0.72	0.78	0.73	0.75	0.73	0.71	0.78	0.72	0.74	0.73	0.70	0.78	0.71
	RF	0.69	0.67	0.65	0.72	0.66	0.70	0.67	0.65	0.74	0.66	0.69	0.67	0.65	0.73	0.66
	SVM	0.74	0.71	0.71	0.76	0.71	0.74	0.73	0.70	0.78	0.71	0.70	0.68	0.67	0.73	0.68
≥50, n = 325	LASSO	0.82	0.69	0.65	0.89	0.67	0.82	0.70	0.65	0.89	0.67	0.81	0.67	0.65	0.87	0.66
	PLS-DA	0.78	0.66	0.45	0.91	0.53	0.78	0.67	0.47	0.91	0.55	0.80	0.69	0.57	0.90	0.62
	RF	0.71	0.49	0.37	0.85	0.42	0.70	0.47	0.36	0.84	0.41	0.73	0.54	0.38	0.87	0.44
	SVM	0.76	0.60	0.53	0.86	0.56	0.76	0.58	0.52	0.85	0.55	0.75	0.57	0.58	0.82	0.58
≥70, n = 182	LASSO	0.86	0.62	0.39	0.95	0.48	0.86	0.61	0.39	0.95	0.47	0.85	0.54	0.40	0.93	0.46
	PLS-DA	0.85	0.63	0.13	0.99	0.22	0.85	0.66	0.14	0.99	0.23	0.84	0.52	0.46	0.98	0.22
	RF	0.81	0.18	0.06	0.95	0.08	0.81	0.19	0.05	0.96	0.08	0.81	0.15	0.04	0.96	0.06
	SVM	0.82	0.36	0.12	0.96	0.17	0.82	0.39	0.17	0.95	0.23	0.83	0.43	0.28	0.93	0.34
Whole plant corn																
≥10, n = 646	LASSO	0.58	0.63	0.66	0.48	0.64	0.58	0.62	0.66	0.48	0.64	0.55	0.60	0.63	0.43	0.61
	PLS-DA	0.53	0.58	0.67	0.35	0.62	0.54	0.58	0.68	0.36	0.63	0.47	0.53	0.62	0.28	0.57
	RF	0.55	0.60	0.64	0.44	0.62	0.55	0.60	0.63	0.43	0.61	0.55	0.60	0.62	0.46	0.61
	SVM	0.55	0.60	0.61	0.47	0.60	0.54	0.60	0.63	0.43	0.61	0.51	0.57	0.59	0.42	0.58
≥30, n = 187	LASSO	0.77	0.14	0.08	0.90	0.10	0.76	0.12	0.07	0.90	0.09	0.78	0.20	0.11	0.91	0.14
	PLS-DA	0.82	0.11	0.01	0.98	0.02	0.82	0.10	0.01	0.98	0.02	0.81	0.04	0.01	0.97	0.01
	RF	0.81	0.10	0.016	0.97	0.03	0.81	0.081	0.02	0.96	0.03	0.82	0.07	0.01	0.97	0.02
	SVM	0.79	0.23	0.11	0.93	0.15	0.79	0.20	0.09	0.93	0.12	0.78	0.16	0.09	0.91	0.11

[1]GBF = grassland-based feed/s; LASSO = least absolute shrinkage and selection operator; PLS-DA = partial least squares discriminant analysis; RF = random forest; SVM = support vector machine.

[2]Or recall.

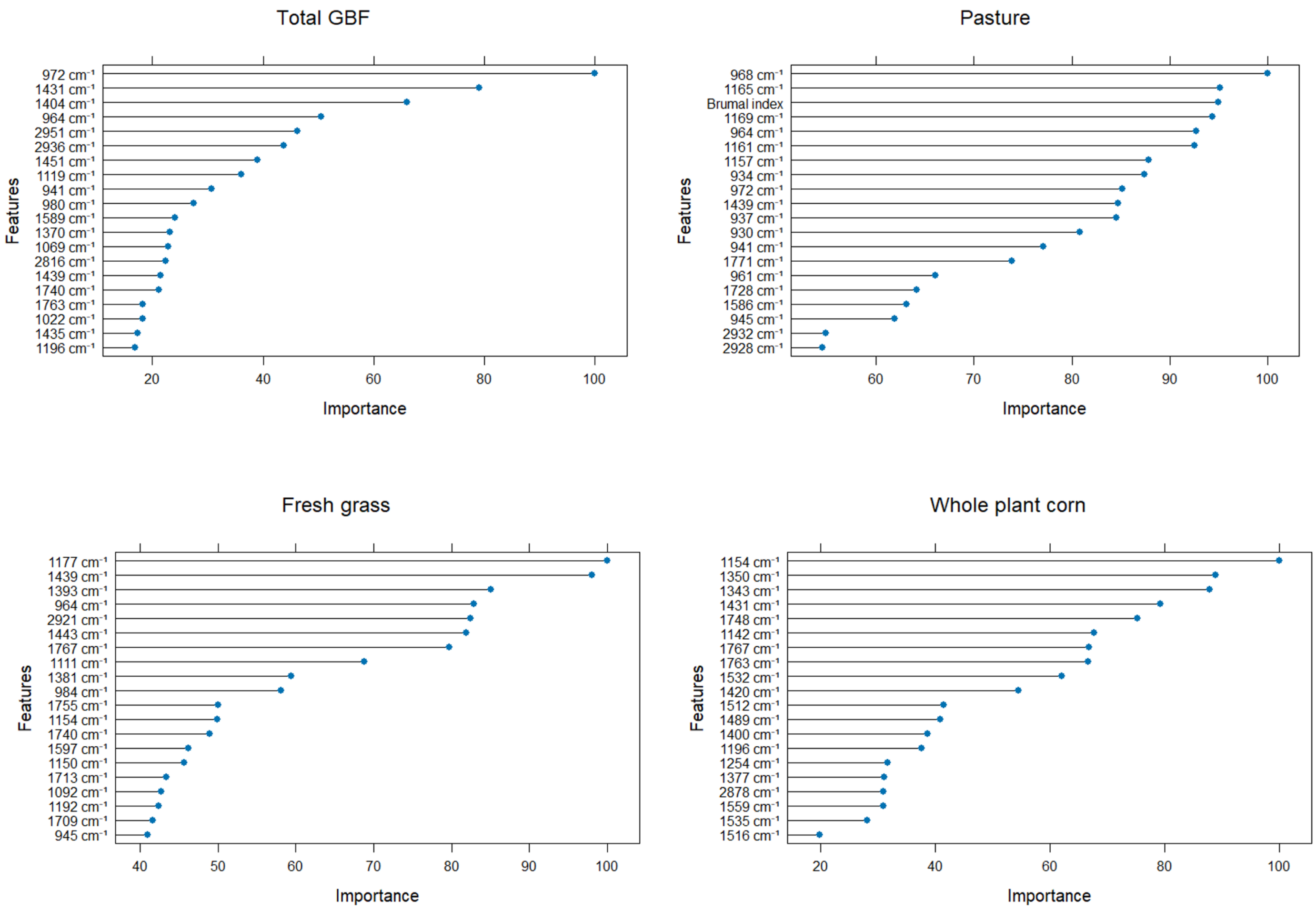

Figure 2. Depiction of the 20 most important covariates (numbers = wavenumbers), in descending order, for the model with the highest precision from each feed group (total grassland-based forages, pasture, fresh herbage [pasture + fresh herbage indoor feeding], and whole plant corn [fresh + conserved]).

considerable. In the PLS-DA model classifying pasture ≥70%, we found was a 5% improvement in the precision of the model; however, a decrease of 11% was recorded in the F1 score of the LASSO model classifying pasture ≥70%. The addition of the breeding association data did not improve model performance greatly; in fact, it decreased model performance in most cases. Dietary pasture proportion ≥20% was better classified than pasture ≥50% or pasture ≥70% (Table 3). Wavenumbers in the C–H bending and C–O stretching (900–1,500/cm) and fat absorption regions (1,736–1,805/cm; 2,855–2,928/cm) of the MIR spectra made up the majority of the 20 most important variables in pasture classification in terms of precision and sensitivity (Figures 2 and 3, Pasture). The Brumal index was one of the top 20 variables in the model with the highest precision, PLS-DA ≥70% with seasonality indices. Predicted fat content was one of the 20 most important variables, in terms of sensitivity, in the LASSO ≥20% model with seasonality indices.

Fresh Herbage (Pasture and Indoors)

Similar to pasture proportion classification, the LASSO models were the most accurate for classifying fresh herbage ≥20%, ≥ 50% and ≥70%, whereas the RF models were the least accurate for classifying all 3 categories of fresh herbage intake. Precision also followed the same trend, where the LASSO models were most precise when classifying fresh herbage intake ≥20% and ≥50%, and the PLS-DA model was most precise when classifying fresh herbage ≥70%. The RF models were the least precise at classifying all 3 categories of fresh herbage intake. This was mirrored in terms of sensitivity, with the RF models performing the worst across all 3 categories of fresh herbage intake and the LASSO models reporting the highest sensitivity across all 3 fresh herbage categories. Specificity was highest in the PLS-DA models classifying fresh herbage ≥50% and ≥70%, and in the LASSO model classifying fresh herbage ≥20%. Specificity was

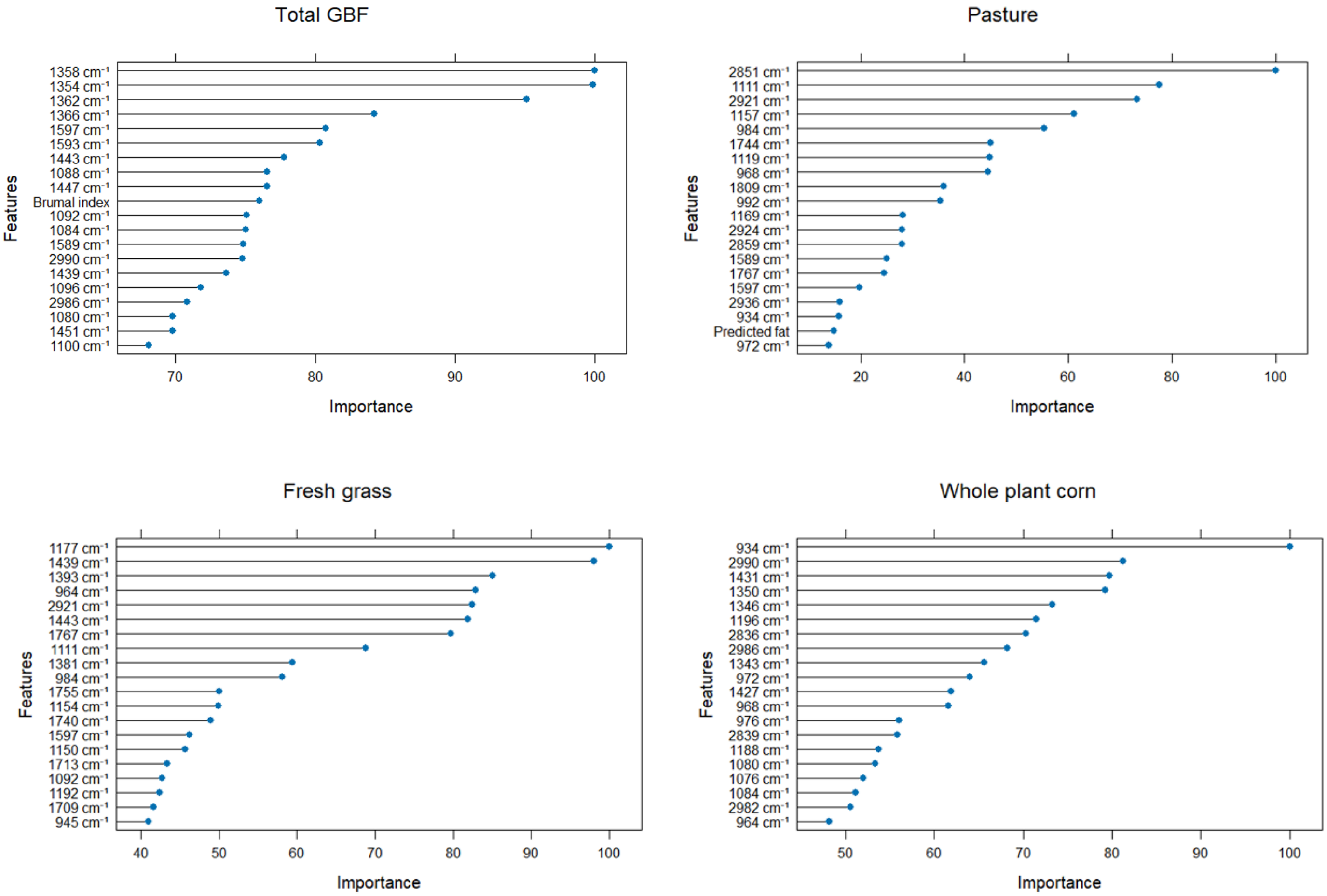

Figure 3. Depiction of the 20 most important covariates (numbers = wavenumbers), in descending order, for the model with the highest sensitivity from each feed group (total grassland-based forages, pasture, fresh herbage [pasture + fresh herbage indoor feeding], and whole plant corn [fresh + conserved]).

lowest in the RF models across all fresh herbage classification categories. Similarly, F1 scores were highest in the LASSO models across all 3 fresh herbage categories and lowest in the RF models across all 3 categories. The addition of the seasonality indices to the models did not garner large changes, and the largest increase of 3% was reported in the PLS-DA model classifying fresh herbage ≥70%. In general, we are only able to report changes of a couple of percentage points with the addition of the breeding association data; however, the PLS-DA model, which improved by 3 percentage points with the addition of the seasonality indices, dropped considerably from the baseline model (covariate set 1). As seen with pasture proportion classification, ≥ 20% fresh herbage in the diet was better classified than fresh herbage ≥50% and ≥70%. Wavenumbers in the C–H bending and C–O stretching (900–1,500/cm) and fat absorption regions (1,736–1,805/cm; 2,855–2,928/cm) of the MIR spectra made up the majority of the 20 most important variables in fresh herbage classification in terms of precision and specificity (Figures 2 and 3, Fresh herbage).

Whole Plant Corn (Fresh and Conserved)

The LASSO model reported the highest accuracy for classifying whole plant corn proportions ≥10%, and for the classification of whole plant corn proportions ≥30%, the PLS-DA model was most accurate. The inverse was reported for the least accurate model, with PLS-DA recording the lowest accuracy for whole plant corn classification ≥10% and the LASSO model recording the lowest accuracy for whole plant corn classification ≥30%. Similarly, the LASSO model classifying whole plant corn proportions ≥10% was the most precise with the SVM model recording the highest precision for classifying whole plant corn proportions ≥30%. The least precise models were recorded as the PLS-DA and SVM models for classifying whole plant corn proportions ≥10% and

≥30%, respectively. In terms of sensitivity, the PLS-DA model was the most sensitive for classifying whole plant corn proportions ≥10%, and the SVM model the most sensitive for classifying whole plant corn proportions ≥30%. The SVM and PLS-DA models were the least sensitive for classifying whole plant corn proportions ≥10% and ≥30%, respectively. Specificity was highest in the LASSO and PLS-DA models for classifying whole plant corn proportions ≥10% and ≥30%, respectively, and lowest in the PS-DA and LASSO models for ≥10% and ≥30% classification, respectively. For F1 scores the LASSO models performed best in both whole plant corn classification categories and the SVM and PLS-DA models performed the most poorly for the classification of whole plant corn proportions ≥10% and ≥30%, respectively. The addition of the seasonality indices barely affected model performance, whereas the addition of the breeding association data seemed to decrease model performance slightly in most cases. Wavenumbers in the C–H bending and C–O stretching (900–1,500/cm) and fat absorption regions (1,736–1,805/cm; 2,855–2,928/cm) of the MIR spectra make up the majority of the 20 most important variables in whole plant corn classification (Figure 2 and 3, Whole plant corn).

DISCUSSION

We tested 4 machine learning algorithms to classify total GBF, pasture, and fresh herbage proportions at herd level and found that the LASSO and PLS-DA models performed best, whereas the RF and SVM models performed less well. The LASSO models performed slightly better than the PLS-DA models in pasture and fresh herbage classification, whereas for total GBF, the LASSO models were marginally better at classifying higher proportions (≥75% and ≥85%) of total GBF and the LASSO and PLS-DA models were very similar at classifying the ≥50% category, with the LASSO model being slightly more specific. However, the RF and SVM models were not far behind the best performing models in each forage category. This provides useful insight as most of the previous research involving GBF binary classification models used PLS-DA (Capuano et al., 2014; Coppa et al., 2021; Soyeurt et al., 2022). We found that the classification of whole plant corn in the diet followed a similar trend, with the LASSO and PLS-DA models generally performing better than the RF and SVM models.

Ideally, GBF prediction models would only include data that is easily available and can be obtained at low costs. For this reason, we assessed the accuracy of our models using only readily available MIR spectral data and milk quality traits (fat and protein contents predicted from MIR spectral data), with or without the inclusion of seasonality indices and breeding association data. As a first attempt to introduce seasonality into our models in a reproducible, nonbiased, consistent manner, we followed a method often used in environmental mapping. Including seasonality in this manner did not greatly affect the classification capabilities of our models; however, in some cases, classification metrics were improved by a couple of percentage points. More advanced seasonality indices are required to include temperature and precipitation into prediction models. However, as a first step, our approach is a low cost, easily available method that requires further testing to allow season to be included consistently in global future prediction models for dietary GBF proportions.

Although the precision and sensitivity improved slightly in some models when the additional breeding association data information was included, it also decreased slightly or remained very similar in other models. Only one of the models we developed and tested (PLS-DA pasture ≥70%) had a decrease of more than 7% in any of the 5 confusion matrix metrics (accuracy, precision, sensitivity, specificity, F1 score) we measured across all 3 sets of covariates for each algorithm. Based on these results, it is not of large relevance if the time costly herd specific data are missing in GBF proportion prediction models.

Although no other study has directly compared statistical approaches for the binary classification of GBF, multiple approaches have been tested by Frizzarin et al. (2021 and 2023) for the classification of complete diets. In their studies, 3 known dietary treatments (grass, grass/clover, and TMR) were discriminated with 11 machine learning methods (Frizzarin et al., 2021). Accuracy ranged from 67% to 97% across methods, with PLS-DA, LASSO, and SVM achieving an accuracy of 97%, 96% and 95%, respectively. This is about 10% higher than the highest accuracy across all models in our study scoring 87%, 88% and 85% for PLS-DA, LASSO, and SVM, respectively. However, these results refer to complete diets and not individual components. When classifying individual components, only PLS-DA has been tested by one other study (Coppa et al., 2021). Our models registered very similar results for total GBF ≥50% but were not as powerful in pasture ≥50% classification. A limitation in our study and a possible reason for the differences in performances is the allocation of 2 sets of MIR spectra to one set of dietary data. Coppa et al. (2021) reported a precision of 89.2% and sensitivity of 84.3% for total GBF ≥50%. In our models, total GBF ≥50% could be classified with a precision of 85% (sensitivity = 90.6%) using the LASSO algorithm with all 3 sets of covariates. This dropped to a precision of 65% (sensitivity = 65%) when classifying total GBF ≥75% a threshold required by the Grassland-Based Meat and Milk production program. This may be because GBF proportions exceeding

70% result in minimal differences in milk composition and are more related to herbage quality rather than quantity (Coppa et al., 2012). Sensitivity was very high (99%) for the PLS-DA algorithm total GBF ≥50% with covariate set 2 (+ seasonality indices), with a corresponding precision of 84% using the same model. This dropped to 59% and 63%, respectively, when increasing total GBF ≥75%. Such high sensitivity leads to a very low specificity (0.06%), which in a real-life scenario could result in overpayments. Pasture ≥70%, relevant for the grazing contribution payment in Switzerland from May to October, could be classified with the PLS-DA algorithm and covariate set 2 (+ seasonality indices) with a precision of 80% but a corresponding sensitivity of 2.6% (accuracy = 87%). The LASSO algorithm for pasture ≥20% with covariate set 2 (+ seasonality indices) registered a precision of 68% and a sensitivity of 70%, satisfying the requirements for the FOAG's Regular Outdoor Exercise program. For pasture ≥50%, Coppa et al. (2021) reported a better precision (82.8%) and sensitivity (78.5%) score than our models for the same proportion of pasture. This may be due to the larger number of MIR spectra (7,607) and farms (1,355) included in their study. For fresh herbage the LASSO algorithm ≥20% with only the MIR spectra (covariate set 1) performed best with a precision and sensitivity of 75% and 76%, respectively. Whole plant corn (negative control) was the most difficult item for our models to classify, in contrast to Coppa et al. (2021), who classified corn specifications for Cantal cheese production with a precision and sensitivity >80%. Regarding our whole plant corn models, precision was highest (63%) with the LASSO algorithm ≥10% and only MIR data (covariate set 1), whereas sensitivity was 68% with the PLS-DA algorithm ≥10% with the seasonality indices. To ensure rigorous predictions, only 1 set of MIR spectra per diet should be included.

CONCLUSIONS

Based on our results, LASSO and PLS-DA appear to be superior to RF and SVM for GBF classification in dairy herd diets. However, performance metrics (accuracy, precision, sensitivity, specificity, F1 score) were often similar across algorithms, rendering the choice of algorithm less critical than expected. Our best-performing model, the LASSO model containing seasonality indices and herd specific information, classified total GBF ≥50% with an accuracy of 78.6%, precision of 85.1%, sensitivity of 90.6%, specificity of 14.1%, and F1 score of 87.7%. However, adding seasonality indices and herd-specific data generally made little difference to the overall model performances. Future research should continue to focus on multifaceted approaches incorporating measured feed proportions, milk MIR spectra, and readily available milk quality traits. Future prediction models may benefit from more developed seasonality indices able to account for temperature and precipitation, rather than only considering sample date as in the present study.

NOTES

This project was funded by Innosuisse (Berne, Switzerland, 32078.1 IP-LS). The authors thank the individuals and institutions that made this study possible. J. Braun (Bern University of Applied Sciences [BFH], School for Agricultural, Forest and Food Sciences [HAFL], Zollikofen, Switzerland), J. Mäder, and N. Matt (BFH, HAFL students) for the collection of the on-farm data; S. Ineichen (BFH, HAFL) for assistance with the milk samples; Braunvieh Schweiz (Zug, Switzerland), Holstein Switzerland (Posieux, Switzerland), and Swissherdbook (Zollikofen, Switzerland) for providing the breeding association data; and Identitas AG (Bern, Switzerland) for supplying the animal traffic data. Our project partners, IP-Suisse (Zollikofen, Switzerland), Suisselab (L. Jungo, U. Birrer, and M. Houmard; Zollikofen, Switzerland), and IG Neue Schweizer Kuh (H. Jörg; Switzerland) are greatly acknowledged. Supplemental material for this article is available at https://doi.org/10.6084/m9.figshare.27095632. Author contributions are as follows: A. Birkinshaw, conceptualization, methodology, software, formal analysis, investigation, resources, data curation, writing (original draft), visualization; M. Sutter, conceptualization, methodology, validation, software, investigation, resources, data curation, writing (review and editing), visualization, project administration; M. Nussbaum, methodology, software, validation, resources, data curation, writing (review and editing); M. Kreuzer: resources, writing (review and editing), visualization, project administration; B. Reidy, conceptualization, validation, resources, writing (review and editing), supervision, project administration, funding acquisition. Approval from the Cantonal Veterinary Office was not required for this study, as there was no contact of any kind (including observations) between the researchers and dairy herds. The authors have not stated any conflicts of interest.

Nonstandard abbreviations used: FOAG = Federal Office of Agriculture; GBF = grassland-based forage; LASSO = least absolute shrinkage and selection operator; MIR = mid-infrared; PLS-DA = partial least squares discriminant analysis; RF = random forest; SVM = support vector machines; TVD = national animal tracing database.

REFERENCES

Agostinelli, C., and U. Lund. 2022. R Package 'circular': Circular Statistics (version 0.4-95). Accessed Dec. 20, 2022. http://r-forge.r-project.org/projects/circular/.

Agroscope. 2016. Swiss Feed Database. Accessed April 23, 2024. https://www.agroscope.admin.ch/agroscope/en/home/services/support-services/feedstuffs/swiss-feed-database.html.

Alothman, M., S. A. Hogan, D. Hennessy, P. Dillon, K. N. Kilcawley, M. O'Donovan, J. Tobin, M. A. Fenelon, and T. F. O'Callaghan. 2019. The grass-fed milk story: Understanding the impact of pasture feeding on the composition and quality of bovine milk. Foods 8:350. https://doi.org/10.3390/foods8080350.

Amatulli, G., S. Domisch, M. N. Tuanmu, B. Parmentier, A. Ranipeta, J. Malczyk, and W. Jetz. 2018. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Sci. Data 5:180040. https://doi.org/10.1038/sdata.2018.40.

Biau, G., and E. Scornet. 2016. A random forest guided tour. Test 25:197–227. https://doi.org/10.1007/s11749-016-0481-7.

Birkinshaw, A., M. Sutter, B. Reidy, L. Jungo, S. Mueller, M. Kreuzer, and M. Terranova. 2023. Evaluation and quantification of associations between commonly suggested milk biomarkers and the proportion of grassland-based feeds in the diets of dairy cows. PLoS One 18:e0282515. https://doi.org/10.1371/journal.pone.0282515.

Bord Bia. 2019. Grass-fed Dairy. Accessed Apr. 4, 2024. https://www.bordbia.ie/farmers-growers/prices-markets/agri-market-insights/grass-fed-dairy-standard/.

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

Brereton, R. G., and G. R. Lloyd. 2014. Partial least squares discriminant analysis: Taking the magic away. J. Chemometr. 28:213–225. https://doi.org/10.1002/cem.2609.

Capuano, E., J. Rademaker, H. van den Bijgaart, and S. M. van Ruth. 2014. Verification of fresh grass feeding, pasture grazing and organic farming by FTIR spectroscopy analysis of bovine milk. Food Res. Int. 60:59–65. https://doi.org/10.1016/j.foodres.2013.12.024.

Chang, J., P. Ciais, T. Gasser, P. Smith, M. Herrero, P. Havlík, M. Obersteiner, B. Guenet, D. S. Goll, W. Li, V. Naipal, S. Peng, C. Qiu, H. Tian, N. Viovy, C. Yue, and D. Zhu. 2021. Climate warming from managed grasslands cancels the cooling effect of carbon sinks in sparsely grazed and natural grasslands. Nat. Commun. 12:118. https://doi.org/10.1038/s41467-020-20406-7.

Coppa, M., B. Martin, C. Agabriel, C. Chassaing, C. Sibra, I. Constant, B. Graulet, and D. Andueza. 2012. Authentication of cow feeding and geographic origin on milk using visible and near-infrared spectroscopy. J. Dairy Sci. 95:5544–5551. https://doi.org/10.3168/jds.2011-5272.

Coppa, M., B. Martin, S. Hulin, J. Guillemin, J. V. Gauzentes, A. Pecou, and D. Andueza. 2021. Prediction of indicators of cow diet composition and authentication of feeding specifications of Protected Designation of Origin cheese using mid-infrared spectroscopy on milk. J. Dairy Sci. 104:112–125. https://doi.org/10.3168/jds.2020-18468.

Dini, I., R. Di Lorenzo, A. Senatore, D. Coppola, and S. Laneri. 2021. Comparison between mid-infrared (ATR-FTIR) spectroscopy and official analysis methods for determination of the concentrations of alcohol, SO_2, and total acids in wine. Separations 8:191. https://doi.org/10.3390/separations8100191.

FOAG (Federal Office of Agriculture). 2019. Animal Tracing Database TVD. Accessed Apr. 4, 2024. https://www.blw.admin.ch/blw/de/home/politik/datenmanagement/agate/tvd.html.

FOAG (Federal Office of Agriculture). 2022. Grassland Ecosystem Services. Accessed Apr. 4, 2024. https://www.agroscope.admin.ch/agroscope/en/home/topics/research-programs/indicate/indicate-projekte/grasland.html.

FOAG (Federal Office of Agriculture). 2023. Tierwohl. Accessed Apr. 4, 2024. https://www.blw.admin.ch/blw/de/home/instrumente/direktzahlungen/produktionssystembeitraege23/tierwohlbeitraege1.html.

FOAG (Federal Office of Agriculture). 2024. Beitrag für graslandbasierte Milch- und Fleischproduktion. Accessed Feb. 28, 2024. https://www.blw.admin.ch/blw/de/home/instrumente/direktzahlungen/produktionssystembeitraege23/beitrag-fuer-graslandbasierte-milch-und-fleischproduktion.html.

Frizzarin, M., T. F. O'Callaghan, T. B. Murphy, D. Hennessy, and A. Casa. 2021. Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets. J. Dairy Sci. 104:12394–12402. https://doi.org/10.3168/jds.2021-20812.

Frizzarin, M., G. Visentin, A. Ferragina, E. Hayes, A. Bevilacqua, B. Dhariyal, K. Domi-jan, H. Khan, G. Ifrim, T. L. Nguyen, J. Meagher, L. Menchetti, A. Singh, S. Whoriskey, R. Williamson, M. Zappaterra, and A. Casa. 2023. Classification of cow diet based on milk Mid Infrared Spectra: A data analysis competition at the "International Workshop on Spectroscopy and Chemometrics 2022." Chemometr. Intell. Lab. Syst. 234:104755. https://doi.org/10.1016/j.chemolab.2023.104755.

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

Hastie, T., and J. Qian. 2006. An introduction to glmnet. Accessed Apr. 4, 2024. https://mran.revolutionanalytics.com/snapshot/2019-12-03/web/packages/glmnet/vignettes/glmnet.pdf.

Hastie, T., R. Tibsharani, and J. Friedman. 2001. Random Forests. Pages 587–603 in The Elements of Statistical Learning. Springer Series in Statistics, Springer, New York, NY.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY.

Hill, L. 2023. The Well-Deserved Hype Behind Organic Grass Milk. Accessed Apr. 4, 2024. https://www.organicvalley.coop/blog/the-well-deserved-hype-behind-organic-grassmilk/.

Infascelli, L., R. Tudisco, P. Iommelli, and F. Capitanio. 2021. Milk quality and animal welfare as a possible marketing lever for the economic development of rural areas in southern Italy. Animals (Basel) 11:1059. https://doi.org/10.3390/ani11041059.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. Support vector machines. Pages 337–368 in An Introduction to Statistical Learning with Applications in R. Springer Science and Business Media, New York, NY.

Jans, F., J. Kessler, A. Münger, and P. Schlegel. 2015. Fütterungsempfehlung für die Milchkuh. Chapter 7 in Fütterungsempfehlung für Wiederkäuer. Eidgenössische Forschungsanstalt für Nutztiere Agroscope (in German). Posieux, Switzerland.

Karatzoglou, A., A. Smola, and K. Hornik. 2022. Kernlab: Kernel-Based Machine Learning Lab. R package version 0.9-31. Apr. 23, 2024. https://cran.r-project.org/web/packages/kernlab/index.html.

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28:1–26. https://doi.org/10.18637/jss.v028.i05.

Kursa, M. B., and W. R. Rudnicki. 2010. Feature selection with the Boruta package. J. Stat. Softw. 36:1–13. https://doi.org/10.18637/jss.v036.i11.

Kursa, M. B. 2022. Package Boruta. Accessed Apr. 23, 2024. https://cran.r-project.org/web/packages/Boruta/Boruta.pdf.

Lee, L. C., C. Y. Liong, and A. A. Jemain. 2018. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps. Analyst 143:3526–3539. https://doi.org/10.1039/C8AN00599K.

Legrand, A. L., M. A. G. von Keyserlingk, and D. M. Weary. 2009. Preference and usage of pasture versus free-stall housing by lactating dairy cattle. J. Dairy Sci. 92:3651–3658. https://doi.org/10.3168/jds.2008-1733.

Liland, K., B. Mevik, and R. Wehrens. 2023. pls: Partial Least Squares and Principal Component Regression. R package version 2.8-2. Apr. 23, 2024. https://CRAN.R-project.org/package=pls.

Marguerite Happy Cow. 2015. Marguerite Happy Cow, Sustainable Dairy Cooperation in the Land of Herve. Accessed Apr. 4, 2024. https://www.margueritehappycow.be/.

Maurice-Van Eijndhoven, M. H. T., H. Soyeurt, F. Dehareng, and M. P. L. Calus. 2013. Validation of fatty acid predictions in milk using mid-infrared spectroscopy across cattle breeds. Animal 7:348–354. https://doi.org/10.1017/S1751731112001218.

Molnar, C. 2022. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Accessed April 23, 2024. https://christophm.github.io/interpretable-ml-book/cite.html.

Moscovici Joubran, A., K. M. Pierce, N. Garvey, L. Shalloo, and T. F. O'Callaghan. 2021. Invited review: A 2020 perspective on pasture-based dairy systems and products. J. Dairy Sci. 104:7364–7382. https://doi.org/10.3168/jds.2020-19776.

Olsen, L. R. 2022. Groupdata2: Creating Groups from Data. R package version 2.0.1. Apr. 23, 2024. https://CRAN.R-project.org/package=groupdata2.

Pereira, C. G., L. C. Luiz, M. J. V. Bell, and V. Anjos. 2020. Near and mid infrared spec-troscopy to assess milk products quality: A review of recent applications. J. Dairy Res. Technol. https://doi.org/10.24966/DRT-9315/100014.

Pérez-Prieto, L. A., and R. Delagarde. 2012. Meta-analysis of the effect of pregrazing pasture mass on pasture intake, milk production, and grazing behavior of dairy cows strip-grazing temperate grasslands. J. Dairy Sci. 95:5317–5330. https://doi.org/10.3168/jds.2012-5609.

R Core Team. 2022. R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Apr. 23, 2024. https://www.R-project.org/.

Rutter, S. M. 2010. Review: Grazing preferences in sheep and cattle—Implications for production, the environment and animal welfare. Can. J. Anim. Sci. 90:285–293. https://doi.org/10.4141/CJAS09119.

Shine, P., and M. D. Murphy. 2022. Over 20 years of machine learning applications on dairy farms: A comprehensive mapping study. Sensors (Basel) 22:52. https://doi.org/10.3390/s22010052.

Signorell, A. 2022. bfsMaps: Plot Maps from Switzerland by Swiss Federal Statistical Office. R package version 0.9.8. Apr. 23, 2024. https://CRAN.R-project.org/package=bfsMaps.

Soyeurt, H., C. Gerards, C. Nickmilder, J. Bindelle, S. Franceschini, F. Dehareng, D. Veselko, C. Bertozzi, N. Gengler, A. Marvuglia, A. Bayram, and A. Tedde. 2022. Prediction of indirect indicators of a grassland-based diet by milk Fourier transform mid-infrared spectroscopy to assess the feeding typologies of dairy farms. Animals (Basel) 12:2663. https://doi.org/10.3390/ani12192663.

Stampa, E., C. Schipmann-Schwarze, and U. Hamm. 2020. Consumer perceptions, preferences, and behavior regarding pasture-raised livestock products: A review. Food Qual. Prefer. 82:103872. https://doi.org/10.1016/j.foodqual.2020.103872.

Tiplady, K. M., T. J. Lopdell, M. D. Littlejohn, and D. J. Garrick. 2020. The evolving role of Fourier-transform mid-infrared spectroscopy in genetic improvement of dairy cattle. J. Anim. Sci. Biotechnol. 11:39. https://doi.org/10.1186/s40104-020-00445-2.

Weidemelk. 2022. What is Meadow Milk? Accessed Apr. 4, 2024. https://www.weidemelk.nl/nl/.

Wickham, H. 2016. Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York, NY. https://ggplot2.tidyverse.org.

Williams, B., C. Halloin, W. Löbel, F. Finklea, E. Lipke, R. Zweigerdt, and S. Cremaschi. 2020. Data driven model development for cardiomyocyte production experimental failure prediction. Pages 1639–1644 in Computer Aided Chemical Engineering. S. Pierucci, F. Manenti, G. L. Bozzano, and D. Manca, ed. Elsevier.

Wright, M. N., and A. Ziegler. 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77:1–17. https://doi.org/10.18637/jss.v077.i01.

ORCIDS

A. Birkinshaw https://orcid.org/0000-0002-0356-7841

M. Sutter https://orcid.org/0000-0003-0314-5697

M. Kreuzer https://orcid.org/0000-0002-9978-1171

B. Reidy https://orcid.org/0000-0002-8619-0209