Application of machine-learning methods to milk mid-infrared spectra for discrimination of cow milk from pasture or total mixed ration diets

M. Frizzarin,[1,2] **T. F. O'Callaghan,**[3,4] **T. B. Murphy,**[1,3] **D. Hennessy,**[2,3] **and A. Casa**[1,3]*

[1]School of Mathematics and Statistics, University College Dublin, Belfield, Dublin 4, Ireland D04 V1W8
[2]Teagasc, Animal & Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland P61 P302
[3]VistaMilk SFI Research Center, Moorepark, Fermoy, Ireland P61 P302
[4]School of Food and Nutritional Sciences, University College Cork, Cork, Ireland T12 Y337

ABSTRACT

The prevalence of "grass-fed" labeled food products on the market has increased in recent years, often commanding a premium price. To date, the majority of methods used for the authentication of grass-fed source products are driven by auditing and inspection of farm records. As such, the ability to verify grass-fed source claims to ensure consumer confidence will be important in the future. Mid-infrared (MIR) spectroscopy is widely used in the dairy industry as a rapid method for the routine monitoring of individual herd milk composition and quality. Further harnessing the data from individual spectra offers a promising and readily implementable strategy to authenticate the milk source at both farm and processor levels. Herein, a comprehensive comparison of the robustness, specificity, and accuracy of 11 machine-learning statistical analysis methods were tested for the discrimination of grass-fed versus non-grass-fed milks based on the MIR spectra of 4,320 milk samples collected from cows on pasture or indoor total mixed ration–based feeding systems over a 3-yr period. Linear discriminant analysis and partial least squares discriminant analysis (PLS-DA) were demonstrated to offer the greatest level of accuracy for the prediction of cow diet from MIR spectra. Parsimonious strategies for the selection of the most discriminating wavelengths within the spectra are also highlighted.
Key words: Fourier-transform mid-infrared spectroscopy, cow diet, food authentication, machine learning

INTRODUCTION

Grass-fed dairy products are becoming more prominent on the market, often demanding a premium price. Their popularity stems from consumer perceptions of enhanced naturalness, nutritional, environmental, and animal-welfare attributes of such products. Many of those claims have some basis in fact as pasture-derived milk products have been demonstrated to possess increased levels of beneficial nutrients such as n-3 PUFA, vaccenic acid, and CLA, while having reduced levels of n-6 fatty acids and palmitic acid (Alothman et al., 2019; Joubran et al., 2021). As a result of this growth in product demand, in some regions, farmers meeting the requirements to be certified as grass-fed milk producers often receive a premium price incentive. In recent years, some countries have implemented their own standards for grass-fed criteria, which is often unique to their region and practices (Joubran et al., 2021). As an example, in Ireland, the Irish grass-fed standard was recently established by Bord Bia, an Irish government organization whose role is promoting Irish agriculture products worldwide through monitoring the production and setting the standards. The Bord Bia grass-fed standard stipulates that for processors to be able to label their products as Bord Bia–verified grass-fed, the milk must average 95% grass-fed on a fresh weight basis, with the minimum acceptable grass-fed figure from individual herds being ≥90% (Bord Bia, 2021). The typical Irish dairy diet consists of 96% forage on a fresh matter basis, which is predominantly grazed pasture plus some grass silage in late autumn and early spring (O'Brien et al., 2018). It is thus expected that 99% of Irish dairy farms will meet the grass-fed threshold (Bord Bia, 2021).

Therefore, there is a need for the development of rapid technologies to distinguish grass-fed from non-grass-fed dairy in an effort to enhance food security and consumer confidence. Several different methods have been demonstrated as being capable of distinguishing between grass-fed and non-grass-fed milk and dairy products including fatty acid profiling by GC coupled with flame ionization detection (O'Callaghan et al., 2017), nuclear magnetic resonance (O'Callaghan et al., 2017), Raman spectroscopy (Gómez-Mascaraque et al., 2020), and mid-infrared spectroscopy (**MIRS**;

Capuano et al., 2014; Klaffenböck et al., 2017). In general, GC coupled with flame ionization detection and nuclear magnetic resonance are offline technologies that require samples to be extracted and prepared before analysis and, as such, are not rapid or high throughput forms of analysis compared with spectroscopic techniques. Both Raman and MIRS technologies use the effect of light that passes through samples to provide information about the components of the samples. The implementation of proper methodologies for the authentication of grass-fed milk and dairy products will be highly beneficial for the future in terms of ensured consumer confidence, auditing purposes, and for validation of manufacturer claims.

The interest in MIRS technology has increased significantly in recent years due to its low-cost, rapid, and nondisruptive analysis process. Mid-infrared spectroscopy is widely used across the dairy industry for the prediction of general macronutrients and other traits such as fat, protein, lactose, and some fatty acids that can form the basis of milk payment to farmers. The knowledge that can be harvested from milk MIRS is ever evolving; for example, MIRS has been used to predict novel milk-related traits such as the coagulation properties of milk (Visentin et al., 2015; El Jabri et al., 2019), individual milk fatty acids (Soyeurt et al., 2006; Bonfatti et al., 2017), as well as animal-related traits such as energy efficiency (McParland et al., 2014), energy intake (McParland and Berry, 2016), and methane emissions (Dehareng et al., 2012). As reported by Capuano et al. (2014) and Klaffenböck et al. (2017), MIRS can also be a useful tool to discriminate between milks from cows fed different diets. Generally, to relate the MIRS spectra to regression and classification traits, partial least squares regression and partial least squares discriminant analysis (**PLS-DA**) are used, respectively. However, other novel statistical tools have produced promising results, showing potential usefulness when analyzing MIRS data (Frizzarin et al., 2021).

The objective of the present study is to gain some insights on the application of novel statistical methods by exploring and comparing their performances for the discrimination of milk from cows fed grass-based diets and cows fed with silages and concentrates. Therefore, in this study, different machine-learning methods, such as ridge regression (**RR**), least absolute shrinkage and selection operator (**LASSO**), elastic net (**EN**), several versions of discriminant analysis, random forest (**RF**), boosting decision trees, principal components linear regression (**PC-LR**), and support vector machine, were considered to further improve the accuracy of predictions of cow diet from MIRS. If highly accurate predictions of cow diet solely based on the MIRS information

were available, it would then be possible to integrate a rapid tool to existing routine analysis that increases the reliability of producer and consumer confidence toward grass-fed dairy products. Last, as a secondary objective of this study, the importance of the individual wavelengths in discriminating between different feeding regimens was examined, which might help further explore the effect of the diet on the specific chemical characteristics of the milk (A. Casa, T. F. O'Callaghan, and T. B. Murphy, unpublished data).

MATERIALS AND METHODS

Data and Data Editing

Data used in this study originated from Teagasc Moorepark Dairy Research Farm (Fermoy, Co. Cork, Ireland) between May and August in 2015, 2016, and 2017. A total of 120 Holstein-Friesian cows from different parities were involved in the experiment across the years, with a mean number of 36 samples per cow, and with some of the cows participating in the experiment in more than 1 yr. Each year, 54 cows were randomly assigned to different dietary treatments for the entire lactation period. The treatment diets included grass (**GRS**), which consisted of cows maintained outdoors on a perennial ryegrass sward only, clover (**CLV**), where cows were maintained outdoors on a perennial ryegrass white clover sward (with an annual average clover content of 20%) only, and TMR, where cows were maintained indoors and fed with a single nutritional mix containing grass silage, maize silage, and concentrates. Further information on the experimental design and dietary treatments have been described by O'Callaghan et al. (2016). The cows were milked twice daily (0730 and 1530 h), and a.m. and p.m. milk samples were collected once weekly from consecutive milkings and analyzed by a Pro-Foss FT6000 (FOSS). A total of 4,364 milk spectra were stored, comprising 1,060 wavelengths in the region from 925 cm^{-1} and 5,010 cm^{-1}.

As a result of the nature of the systems and strong similarities from a compositional perspective, GRS and CLV classes have been merged together, thus originating a general pasture-based diet group. In fact, as a preliminary exploration of the data, partially acting as the motivation for the subsequent food authenticity analyses reported in the next sections, a linear discriminant analysis was been considered. Figure 1 provides a graphical illustration of the results obtained where the scatter plot of the spectra projections in the latent space spanned by the 2 discriminant functions is shown. This clearly shows how the GRS and CLV classes tend
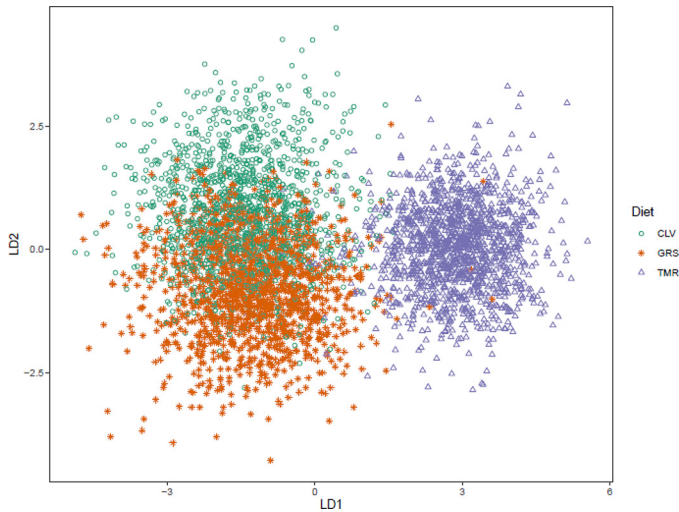
Figure 1. Scatter plot of the observed spectra projected on the space spanned by the two latent variables (LD1 and LD2) associated with the discriminant functions for milk from cows fed grass-only (GRS), grass-white clover (CLV), and a TMR from May to August in 2015, 2016, and 2017. Different colors indicate different diet regimens.

to be more similar, whereas the spectral data seem to contain sufficient information to properly discriminate the pasture and the TMR groups.

For the purpose of this study only, milk samples produced between May and August (representing summer milk) each year were used for the analyses. The final data set consisted of 2,931 milk samples from pasture-fed animals and 1,389 from TMR-fed animals. The spectra were provided in their transmittance values; thus, they have been transformed to absorbance by taking the $\log_{10}$ of the reciprocal of the transmittance value. The Mahalanobis distances from the center have been calculated for all the spectra using their first 8 principal components (**PC**), as these explained 90% of the variability, and outlier spectra were deleted if their corresponding Mahalanobis distance from the center was greater than the 0.99 quantile of all the computed distances. After outlier removal, the data considered in the analyses comprised 4,261 spectra. The regions between 1,720 and 1,592 cm^{-1}, between 3,698 and 2,996 cm^{-1}, and >3,818 cm^{-1} were discarded as high-noise-level regions, resulting in 533 wavelengths used for the analyses. The final data set was then divided into training (60%, n = 2,502) and external validation data sets (40%, n = 1,759) using the groupdata2 package (Olsen, 2020). To avoid potential overfitting issues, when multiple observations were available for one animal, all these observations were included either in the training or in the external validation data set.

All the analyses of the present paper were conducted in the R environment (https://www.r-project.org/).

The code and the data are publicly available at https://github.com/AlessandroCasa/DietDiscriminationProj.

Statistical Analyses

Eleven different classification methods were tested to classify the cow diet from milk spectra. These methods have been grouped according to their characteristics into the following 4 distinct groups: regularized logistic regression methods, discriminant analysis methods, ensemble methods, and other methods. For a general review of many of these techniques, readers may refer to Hastie et al. (2001). Some of these approaches require a careful tuning of parameters, possibly influencing the resulting performances. In the analyses conducted herein, these parameters have been typically selected by performing cross-validation on the training test unless otherwise specified.

Regularized Logistic Regression Methods. Regression models were estimated by optimizing a function that includes a penalty term that shrinks the estimates of the regression coefficients toward zero. This approach, also referred to as regularization, induces a bias in the estimated parameters, but, at the same time, it reduces their variances. This possibly improves the general fit of the model, especially in settings such as the one faced in this work, where a large number of highly correlated variables have been measured. Although widely used in a linear regression framework, these approaches can be adopted, even with classification purposes in mind, by considering penalized versions of logistic regression. In the following, we consider 3 different techniques, namely RR (Hoerl and Kennard, 1970), LASSO (Tibshirani, 1996), and EN (Zou and Hastie, 2005). The RR shrinks the parameters toward zero, but it does not yield estimates that are exactly equal to zero; therefore, all the wavelengths were included in the final model. On the other hand, the different penalization term in the LASSO allows regression coefficients to be equal to zero, and thus it automatically performs variable selection and leads to a gain in terms of interpretability. Last, EN represents a compromise between RR and LASSO, incorporating the main features of both the methods. On the one hand, this method tends to shrink the coefficients of correlated wavelengths toward each other, whereas on the other, it still provides indications about which wavelengths are useless for discrimination purposes.

Finally, because all these approaches include penalization terms in the estimation procedure, the selection of 1 or more tuning parameters, which govern the strength of the penalization itself, is required. These parameters must be carefully selected because they may strongly affect the results. The package glmnet (Friedman et al.,

2010) was used when considering these techniques in the following analyses.

Discriminant Analysis Methods. Discriminant analysis provides a proper approach to classification where each different class in the data is described by a distinct statistical distribution. More specifically, discriminant analysis strategies find combinations of the original variables that better separate the classes observed in the data. These combinations are called discriminant functions and they can subsequently be used as classifiers. Different strategies have been considered in the present work, such as linear discriminant analysis (**LDA**), model-based discriminant analysis (**MB-DA**; Bensmail and Celeux, 1996), PLS-DA, and variable-selection LDA (**VarSel-DA**; Murphy et al., 2010). In the first method, the behavior in the different classes is assumed to be described via Gaussian distributions with the same covariance among the observed variables. As a consequence, the discriminative functions are a linear combination of the wavelengths. Model-based discriminant analysis proposes a regularized modification of such a method, which is particularly useful when dealing with a high number of variables, by considering specific parsimony-inducing parametrizations of the class-specific covariance matrices. The PLS-DA performs discriminant analysis that considers new variables (i.e., factors) that are linear combinations of the original wavelengths. The new variables are constructed via partial least squares (PLS; Garthwaite, 1994), which is a dimension reduction method that builds the new features by exploiting both the spectral information and the information about the feeding regimen, thus detecting directions in the data space that best explain both. Last, VarSel-DA implements a procedure, originally introduced for the food authenticity framework, where the discriminant functions are built using a carefully chosen subset of wavelengths deemed to be the most relevant for classification purposes.

The R packages used to implement the mentioned approaches are MASS (Venables and Ripley, 2002) for LDA, mclust (Scrucca et al., 2016) for MB-DA, and caret (Kuhn, 2020) for PLS-DA, whereas VarSel-DA is not included in an R package and has been implemented by one of the authors.

Ensemble Methods. The rationale behind these approaches consists of combining the strengths of a collection of base classifiers while simultaneously relieving their limits. In the present study, we considered RF (Breiman, 2001) and boosting (Freund and Schapire, 1996); both of these approaches can be applied to improve many different methods. In this work, we considered classification trees (Breiman et al., 1984) as base classifiers. Random forests produce different predictions from multiple classification trees and then combines them into a final consensus classification. Classification tree methods partition the predictor space recursively into simple regions, beginning from a root node and creating branches determined by different splitting rules based on the values of the predictors (here, the wavelengths). The terminal nodes of the tree, also called the leaves, represent class labels and are subsequently used to classify the spectra. Classification trees often produce low bias but highly variable predictions; however, by averaging predictions from many trees and considering a random sample of wavelengths when defining the splitting rules creating different branches, RF often overcome this issue and produce more accurate and stable results.

Boosting, different from RF, grows different trees sequentially on modified versions of the original data set. More specifically, at each stage of the algorithm, a tree is fitted to the residual obtained from the previously built tree. This enables improvement of the model fit and prediction in those areas of the predictor space where the performances of a single classification tree were poor.

The R package used to implement RF was randomForest (Liaw and Wiener, 2002). The number of trees to grow was set to 1,000, and $p^{1/2}$ wavelengths were sampled as candidates at each split, with p being the total number of wavelengths; both of these parameters were user specified. For the boosting approach, the package ada (Culp et al., 2016) was used. The number of boosting iterations was fixed to 500, whereas the default settings were used for the other regularization parameters.

Other Methods. Some other methodologies, not falling into the categories outlined above, were used in this work. These include PC-LR, where a logistic regression model is fitted using the PC scores derived from the spectra as covariates instead of the original wavelengths. The PC, similar to what is done in the PLS framework, are built by detecting the directions in the data space that best explain the variability in the observed spectra. The major difference with respect to PLS is due to the fact that the relationship between the spectra and the information on the feeding treatments is not taken into account. In the analysis, the number of PC has to be carefully chosen; in this study, we selected it so that the newly built variables retained 95% of the variability of the original data.

Last, support vector machine (Cristianini and Shawe-Taylor, 2000) was considered. This method enlarges the feature space using kernels and searches for linear boundaries among the classes in this space. In doing so, it easily allows the building of nonlinear boundaries in the original wavelength space, thus possibly resulting in a more flexible classifier. The R package used was

e1071 (Meyer et al., 2020), in which a radial kernel was used.

Measures of Classification Performance

The performance of all the classification tools mentioned above were evaluated considering different measures. In particular, the sensitivity (i.e., the proportion of samples from pasture diet correctly identified), the specificity (i.e., the proportion of samples from TMR diet correctly identified), and the accuracy (i.e., the overall proportion of samples correctly identified) were considered. Moreover, the F1 score was also considered. The F1 score is defined as $F1 = 2TP/(2TP + FP + FN)$, where TP = true positive, FP = false positive, and FN = false negative, and where the pasture diet is treated as the positive class. This measure takes a value equal to 1 in the case of a perfect classification and a value equal to 0 when no samples coming from the pasture class are classified correctly. Last, we computed the Cohen's kappa coefficient, which is defined, in the case of binary confusion matrices, as follows:

$$\kappa = 2(TP \times TN - FN \times FP)/[(TP + FP)(FP + TN) + (TP + FN)(FN + TN)],$$

where TN = true negative. This coefficient accounts for possible agreement by chance between the true and the predicted diets, and it might be particularly useful in applications with imbalanced classes. It takes values from -1 to 1, with 1 indicating complete agreement, 0 by chance agreement, and negative values showing that the agreement is worse than random. All the measures were computed on the external validation set to have a plausible evaluation of the generalization error for all the considered methods.

RESULTS

Classification Results

Table 1 summarizes the accuracy, sensitivity, specificity, F1 score, and Cohen's kappa coefficient of the external validation set for all the methods tested. Linear discriminant analyses and PLS-DA had the greatest level of accuracy (0.968) as well as the greatest F1 score (0.975). Linear discriminant analysis also had the greatest specificity (0.980), whereas PLS-DA had the greatest sensitivity (0.962). The lowest accuracy was achieved by PC-LR (0.667), as well as the lowest specificity (0.117) and the second lowest F1 score (0.790). Random forest had the lowest F1 score (0.781) and also had the lowest sensitivity (0.827). Specificity had greater variability than sensitivity, with a range of 0.863 and 0.135, respectively. Accuracy had greater variability with respect to the F1 score (range of 0.301 and 0.194, respectively). The Cohen's kappa coefficients values had a higher variability with respect to the other measures considered, showing that the performance of some of the methods (e.g., PC-LR, RF and boosting) were highly influenced by the unbalancedness in the number of observations belonging to the 2 classes.

Wavelength Selection

Figure 2 provides a graphical illustration of the wavelengths selected as more influential, according to the different methods, in discriminating milk samples from cows fed different diets. Some of the methods considered, such as LASSO, EN, and VarSel-DA, automatically perform variable selection in their estimation procedure, thus providing a proper indication on the wavelengths deemed as more relevant. On the other hand, PLS-DA, RF, and boosting use all the observed

Table 1. Summary of the classification methods with the respective accuracy, sensitivity, specificity, F1 score, and Cohen's kappa from the external validation set

Method[1]	Accuracy	Sensitivity	Specificity	F1 score	Cohen's kappa
RR	0.880	0.933	0.779	0.911	0.730
LASSO	0.959	0.953	0.970	0.968	0.910
EN	0.951	0.946	0.960	0.962	0.890
LDA	0.968	0.961	0.980	0.975	0.930
MB-DA	0.964	0.959	0.972	0.972	0.920
PLS-DA	0.968	0.962	0.977	0.975	0.930
VarSel-DA	0.890	0.913	0.845	0.916	0.760
RF	0.696	0.827	0.447	0.781	0.290
Boosting	0.754	0.842	0.587	0.818	0.440
PC-LR	0.667	0.956	0.117	0.790	0.090
SVM	0.947	0.938	0.962	0.958	0.880

[1]RR = ridge regression; LASSO = least absolute shrinkage and selection operator; EN = elastic net; LDA = linear discriminant analysis; MB-DA = model-based discriminant analysis; PLS-DA = partial least squares discriminant analysis; VarSel-DA = variable-selection discriminant analysis; RF = random forest; PC-LR = principal components linear regression; SVM = support vector machine.

variables to build their predictions. Nonetheless, these methodologies provide some measures of variable importance that have been considered to provide some insights about the most influential wavelengths. More specifically, the wavelengths were ordered according to their variable importance scores, and the ones in the upper quartile (i.e., the 75th percentile) were selected. As a consequence, the number of variables displayed in Figure 2 for PLS-DA, RF, and boosting are the same across the methods and are equal to 134. In contrast, LASSO and EN selected 140 and 503 wavelengths, respectively, whereas VarSel-DA was the most parsimonious, choosing a subset of 14 variables.

A post hoc analysis showed how some wavelengths tended to be selected as important from more than one of the methods considered, providing robust indications about which spectral regions are more influenced by the feeding regimen. In particular, the region spanning from 964.23 cm^{-1} to $1,469.49$ cm^{-1} seemed to be particularly relevant, with 3 wavelengths in this interval being selected by all the methods (i.e., 968.09 cm^{-1}, $1,438.63$ cm^{-1}, and $1,465.63$ cm^{-1}).

DISCUSSION

The accurate prediction of cow diet is fundamental to support a system that promotes grass-fed products. In fact, in a market where grass-fed products can command greater prices, the origin of the products must be ensured, avoiding fraud. It is thus necessary to authenticate the grass-fed products in a rapid and low-cost manner. Mid-infrared spectroscopy, combined with appropriate statistical tools, can provide these results.

Classification of the Diet

The suitability of MIRS spectra to predict cow diet has been tested by Klaffenböck et al. (2017) and Capuano et al. (2014), but with different purposes. Klaffenböck et al. (2017) used PLS regression to predict
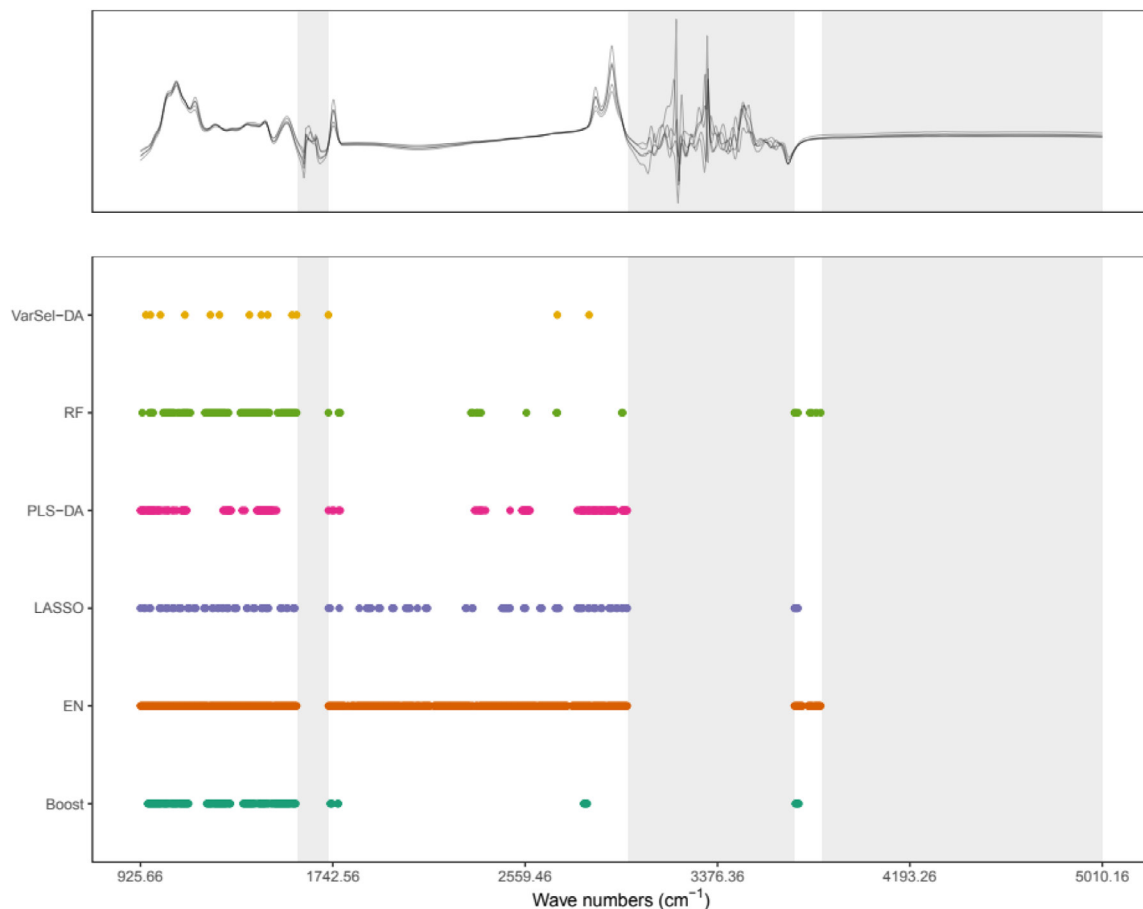

Figure 2. Top panel: subset of observed mid-infrared milk spectra. Bottom panel: representation of the wavelengths selected as more relevant by different methods as listed on the y-axis. The gray-shaded areas represent the highly noisy and water-related wavelength regions, which have been removed from the analyses. VarSel-DA = variable-selection linear discriminant analysis; RF = random forest; PLS-DA = partial least squares discriminant analysis; LASSO = least absolute shrinkage and selection operator; EN = elastic net; Boost = boosting.

the components of the cow diets, whereas Capuano et al. (2014) used PLS-DA to discriminate between milk from fresh grass feeding, pasture grazing, and organic farming. Results from Klaffenböck et al. (2017) showed a range in R^2 values from 0.35 to 0.76 (for hay expressed as percentages of feedstuffs in the total feed ration and pasture expressed as percentages of feedstuffs in the total feed ration, respectively). Thus, pasture has been the variable better predicted by the model. Capuano et al. (2014) reported sensitivity and specificity of 88% and 83% for milk from cows that had fresh grass in the daily ration and milk from cows that did not have fresh grass, respectively. Still, the results of these papers show the potential of MIRS to predict cow diet.

In the present paper, the discriminant analyses methods such as LDA, PLS-DA, and MB-DA provided the best results with an accuracy greater than 0.960. Specificity had the greatest range in the variability of the results compared with sensitivity (i.e., 0.863 and 0.135, respectively), even if specificity also had the greatest result compared with sensitivity (i.e., 0.980 and 0.962, respectively). These results highlight the tendency of the best methods (LDA and PLS-DA) to correctly classify samples from a TMR diet. All methods were able to provide accurate results on classifying pasture derived milks (lowest sensitivity reported by RF, and equal to 0.827). These results can be partly explained by the greatest number of observations belonging to pasture; therefore, all the models were able to recognize this type of diet. These insights were confirmed by the values of the Cohen's kappa coefficients. The analysis of this measure, jointly with the observation of the sensibility and specificity values, showed how some methods (PC-LR, RF, and boosting) produce satisfactory results in terms of accuracy, mainly because of the unbalancedness between the 2 classes.

In the analyses reported in this work, the best performing methods, according to all the measures of classification performance, were the discriminant analysis ones. According to Hastie et al. (2001), a reason motivating these performances might be that "the data can only support simple decision boundaries such as linear or quadratic," advocating the use of simpler models that can be more reliably estimated with respect to more complex alternatives. In fact, even if the data are very high dimensional, the strong correlations among the variables imply that they lie in a lower dimensional manifold, which appears to be adequately captured and described by discriminant analysis techniques. Moreover, these approaches belong to the class of the generative methods [as opposed to the discriminative ones, see Bishop and Lasserre (2007), for a discussion], which have been shown to "present several advantages

in terms of modelling and understanding" when dealing with spectroscopic data (Jacques et al., 2010).

The good performance obtained when considering regularized logistic regression approaches might be due to the similarity between LDA and the logistic regression itself. In fact, as highlighted in section 4.4.5 of the book by Hastie et al. (2001), the 2 approaches share some strong connections, and they often provide similar results in practical applications.

Last, PC-LR provided the worst classification performances, with a Cohen's kappa coefficient indicating an almost completely random agreement. Despite being widely used, principal component analysis might be harmful if used to reduce the number of variables when the final aim is to discriminate between different classes. In fact, as stated above and as opposed to PLS-DA, when building the PC, the relationships between the different diets and the observed spectra are not taken into account. Therefore, as stated in Chang (1983), principal component analysis may hide the group structure, thus providing poor results in terms of classification.

There is an increasing demand for grass-fed milk. Developing rapid methods such as MIRS to determine the origin of the milks in terms of cow diet will be important in substantiating any marketing claims in terms of true grass- or pasture-fed milk. The results from the present study offer a promising insight and can be easily and successfully integrated and used by milk manufactures to identify or authenticate grass-fed milk. The success of the discrimination between cow diet is already evident by looking at the preliminary analyses (i.e., Figure 1). The 2 classes are clearly distinguishable, confirming that the data contains adequate information to discriminate between milk samples from cows fed different diets. Again, the graphical representation has been obtained considering discriminant analysis tools, whereas, for example, discrimination among diets was not visible in the space spanned by the first PC. When considering that grass and clover diets were both pasture-based systems, the similarities between the feeding regimens were reflected in the narrower separation between the points.

The experiment designed by O'Callaghan et al. (2016) constructs a controlled experimental setting. Nonetheless, it represents the first study of its kind in Ireland, and probably the world, to experimentally compare controlled pasture versus TMR diets across a full lactation. As such, it can be seen as a starting point for future works to explore the relation between the cow feeding regimens and quality of milk in less controlled experimental environments. For example, although the analyses were conducted using individual

animal milk samples, it would have been interesting to also test the discriminative ability of the methods on bulk milks from TMR-fed and pasture-fed herds. In this case, a possible source of variation, due to animal specific features, would be avoided. Therefore, we think that when the farming system is kept constant across all cows on the same herd, the calibrations developed should be able to perform well in that scenario as well.

Wavelength Selection

In their work, Bittante and Cecchinato (2013) reported the relationship between the different chemical bonds and the MIRS wavelengths. Based on the important wavelengths selected by the different methods (such as 968.09 cm^{-1}, 1438.63 cm^{-1}, and 1,465.63 cm^{-1}), chemical bonds such as trans-disubstituted alkenes (associated with wavelength 965 cm^{-1}), the double bond C=C (associated with wavelength 1,450 cm^{-1}), and the methylene (associated with wavelength 1,470 cm^{-1}) were particularly relevant for discriminating among cow diets. Other components associated with important wavelengths identified by the different methods were alkyl bonds, monosubstituted alkenes bond, the C–O bonds (which includes alcohols, phenols, ethers, carboxylic acids, and esters), the fluoroalkanes bond, and the N–O bonds (nitro compounds, which includes aliphatic and aromatic bonds). In recent years, MIRS has been successfully used to predict concentrations of fatty acids in milk (Soyeurt et al., 2006; Bonfatti et al., 2017), which can be incorporated into the routine weekly analysis of milks and give important insights into herd health (de novo fatty acids) and prediction of product properties such as butter texture (saturated fatty acids).

Some of the methods tended to select wavelengths that were quite close to the regions associated with water. This can be associated with the presence of information also on the regions that is usually considered to be highly noisy and thus uninformative. It might be worth further exploring this phenomenon, as it would be interesting to evaluate if these regions unexpectedly carry useful information that can somehow be considered during the analyses.

The analysis of the important wavelengths can be useful to develop MIRS instruments that use only specific wavelengths, producing cheaper instruments focused on the prediction of cow diet. Variable-selection discriminant analysis identified 14 wavelengths, predicting cow diet with an accuracy of 0.89. This instrument could be used both in the manufacturing industry, to test the milk in the different process stages, and at the supermarket, to test some random samples.

CONCLUSIONS

This research provided a comprehensive overview on the ability of several statistical and machine-learning techniques to distinguish between pasture-based and nonpasture-derived milk products. This application might be widely useful, adding value to the international dairy industry as the demand for grass-fed dairy continues to rise. Linear discriminant analysis and PLS-DA offered the greatest level of accuracy for the prediction of cow diet from MIRS. However, further work expanding the sample database in terms of diet and stage of lactation would be beneficial to improve robustness, mainly due to the difficulties in discriminating cow diet in pasture-based systems in early lactation where concentrates are typically fed to cows to avoid severe negative energy balance. The collection of MIRS from individual farms is already routine on a lot of dairy regions, forming the basis for farmer payment schemes. As such, the implementation of such models to further harness information for these spectra would be beneficial to the industry across the supply chain. Furthermore, our understanding of the factors affecting or represented by the individual wavelengths of MIRS is ever expanding. Herein, wavelengths that responded to animal diet are outlined, giving greater insight into intrinsic factors affecting the MIRS of milk.

REFERENCES

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.

Bensmail, H., and G. Celeux. 1996. Regularized Gaussian discriminant analysis through eigenvalue decomposition. J. Am. Stat. Assoc. 91:1743–1748. https://doi.org/10.1080/01621459.1996.10476746.

Bishop, C. M., and J. Lasserre. 2007. Generative or discriminative? Getting the best of both worlds. Bayesian Statistics 8:3–24.

Bittante, G., and A. Cecchinato. 2013. Genetic analysis of the Fourier-transform infrared spectra of bovine milk with emphasis on individual wavelengths related to specific chemical bonds. J. Dairy Sci. 96:5991–6006. https://doi.org/10.3168/jds.2013-6583.

Bord Bia (Irish food board). 2021. Grass-Fed Dairy Standard. Accessed May 15, 2021. https://www.bordbia.ie/farmers-growers/prices-markets/agri-market-insights/grass-fed-dairy-standard/.

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

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

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

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, W. 1983. On using principal components before separating a mixture of two multivariate normal distributions. J. R. Stat. Soc. Ser. C Appl. Stat. 32:267–275. https://doi.org/10.2307/2347949.

Cristianini, N., and J. Shawe-Taylor. 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press.

Culp, M., K. Johnson, and G. Michailidis. 2016. ada: The R Package Ada for Stochastic Boosting. R package version 2.0–5. https://CRAN.R-project.org/package=ada

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

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

Freund, Y., and R. Schapire. 1996. Experiments with a new boosting algorithm. Pages 148–156 in Machine Learning: Proceedings of the Thirteenth International Conference. Morgan Kaufman.

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

Frizzarin, M., I. C. Gormley, D. P. Berry, T. B. Murphy, A. Casa, A. Lynch, and S. McParland. 2021. Predicting cow milk quality traits from routinely available milk spectra using statistical machine learning methods. J. Dairy Sci. 104:7438–7447. https://doi.org/10.3168/jds.2020-19576.

Garthwaite, P. H. 1994. An interpretation of partial least squares. J. Am. Stat. Assoc. 89:122–127. https://doi.org/10.1080/01621459.1994.10476452.

Gómez-Mascaraque, L. G., K. Kilcawley, D. Hennessy, J. T. Tobin, and T. F. O'Callaghan. 2020. Raman spectroscopy: A rapid method to assess the effects of pasture feeding on the nutritional quality of butter. J. Dairy Sci. 103:8721–8731. https://doi.org/10.3168/jds.2020-18716.

Hastie, T., R. Tibshirani, and J. Friedman. 2001. The Elements of Statistical Learning. Springer.

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

Jacques, J., C. Bouveyron, S. Girard, O. Devos, L. Duponchel, and C. Ruckebusch. 2010. Gaussian mixture models for the classification of high-dimensional vibrational spectroscopy data. J. Chemometr. 24:719–727. https://doi.org/10.1002/cem.1355.

Joubran, A. M., 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.

Klaffenböck, M., A. Steinwidder, C. Fasching, G. Terler, L. Gruber, G. Mészáros, and J. Sölkner. 2017. The use of mid-infrared spectrometry to estimate the ration composition of lactating dairy cows. J. Dairy Sci. 100:5411–5421. https://doi.org/10.3168/jds.2016-12189.

Kuhn, M. 2020. caret: Classification and Regression Training. R package version 6.0–85. https://CRAN.R-project.org/package=caret.

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

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

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

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2020. e1071: Misc Functions of the Department of Statistics, Probability Theory Group, TU Wien. R package version 1.7–4. https://CRAN.R-project.org/package=e1071

Murphy, T. B., N. Dean, and A. E. Raftery. 2010. Variable selection and updating in model-based discriminant analysis for high-dimensional data with food authenticity applications. Ann. Appl. Stat. 4:396–421. https://doi.org/10.1214/09-AOAS279.

O'Brien, D., B. Moran, and L. Shalloo. 2018. A national methodology to quantify the diet of grazing dairy cows. J. Dairy Sci. 101:8595–8604. https://doi.org/10.3168/jds.2017-13604.

O'Callaghan, T. F., D. Hennessy, S. McAuliffe, K. N. Kilcawley, M. O'Donovan, P. Dillon, R. P. Ross, and C. Stanton. 2016. Effect of pasture versus indoor feeding systems on raw milk composition and quality over an entire lactation. J. Dairy Sci. 99:9424–9440. https://doi.org/10.3168/jds.2016-10985.

O'Callaghan, T. F., D. T. Mannion, D. Hennessy, S. Mcauliffe, M. G. O'Sullivan, N. Leeuwendaal, T. P. Beresford, P. Dillon, K. N. Kilcawley, J. J. Sheehan, R. P. Ross, and C. Stanton. 2017. Effect of pasture versus indoor feeding systems on quality characteristics, nutritional composition, and sensory and volatile properties of full-fat Cheddar cheese. J. Dairy Sci. 100:6053–6073. https://doi.org/10.3168/jds.2016-12508.

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

Scrucca, L., M. Fop, T. B. Murphy, and A. E. Raftery. 2016. mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. R J. 8:289–317. https://doi.org/10.32614/RJ-2016-021.

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

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

Venables, W. N., and B. D. Ripley. 2002. Modern Applied Statistics with S. 4th ed. Springer.

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

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

ORCIDS

M. Frizzarin ● https://orcid.org/0000-0001-7608-5504

T. F. O'Callaghan ● https://orcid.org/0000-0003-2684-7253

T. B. Murphy ● https://orcid.org/0000-0002-5668-7046

D. Hennessy ● https://orcid.org/0000-0002-7375-3754

A. Casa ● https://orcid.org/0000-0002-2929-3850