Journal of Food Composition and Analysis 138 (2025) 107019

Contents lists available at ScienceDirect

Journal of Food Composition and Analysis

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

Classification of "Ricotta" whey cheese from different milk and Designation of Origin-protected samples through infrared spectroscopy and chemometric analysis

Martina Foschi, Alessandra Biancolillo *, Samantha Reale, Francesco Poles, Angelo Antonio D'Archivio

Department of Physical and Chemical Sciences, University of L'Aquila, Via Vetoio, Coppito, L'Aquila 67100, Italy

ARTICLE INFO

Keywords:
Ricotta whey cheese
FT-IR
Classification
Dairy
SIMCA
PDO cheese
Discriminant analysis
Class-modelling

ABSTRACT

Whey cheeses are produced in various parts of the world, such as Portugal, Spain, and Turkey. In Italy, whey cheese goes under the name "ricotta". This study investigates the classification of ricotta whey cheese derived from various milk sources (either protected designation of origin (PDO) or not) using an Attenuated Total Reflectance Fourier-transform infrared (ATR-FTIR) spectroscopy combined with chemometric analysis. Employing the SPORT-LDA method, which can incorporate Variable Importance in Projection (VIP) analysis, 287 samples of ricotta cheese produced using milk from four different animals (sheep, cow, goat, and water buffalo) were classified according to the animal origin. This led to the correct classification of 97 % of the test samples (3 misclassified samples over 97). VIP analysis revealed that the spectral ranges of 3300–3100 cm^{-1}, 2900–2800 cm^{-1}, and 1700–1300 cm^{-1} are consistently relevant across all milk sources, thanks to the key molecular vibrations associated with protein structures, lipid content, and water. Eventually, the analysis was circumscribed to sheep ricotta cheeses, because some of these present the PDO quality mark. SIMCA was used to classify PDO samples with respect to the Non-PDO sheep ricotta individuals. The application of SIMCA to model class PDO led to 82.1 % of sensitivity and 82.7 % of specificity (in external validation). The findings underscore the robustness of ATR-FTIR spectroscopy and chemometrics in maintaining the integrity of PDO products and ensuring quality control.

1. Introduction

Ricotta whey cheese, a unique dairy product with deep-rooted culinary traditions, holds a distinctive place in the realm of artisanal and regional cheese-making. Originating from the liquid byproduct, or whey, left behind after the enzymatic coagulation of milk during cheese production, ricotta whey cheese is an exemplar of resourcefulness in the utilization of dairy industry byproducts. This cheese variant is crafted using traditional methods in various parts of the world, such as Italy, Portugal, Spain, and Turkey, each region contributing its nuances to the production process.

Understanding the importance of ricotta whey cheese involves recognizing its role in sustainable practices within the dairy industry. By repurposing whey, a byproduct that often poses environmental challenges due to its disposal, the production of ricotta whey cheese contributes to minimizing waste and fostering a more sustainable approach to dairy processing.

The term "ricotta," translating to "cooked again" in Italian, reflects the method by which this dairy product is meticulously fashioned. Despite not being categorized as a true cheese under Italian law, ricotta stands out as a versatile dairy creation. Its production involves the application of heat-induced coagulation, typically at temperatures ranging from 85°C - 90 °C. Acidifying agents, commonly lemon or vinegar, are introduced to the whey, initiating the formation of a curd composed of whey proteins, namely albumin and globulin. The resulting curd, characterized by a modest consistency, encapsulates the fats derived from the whey. The curd is then separated, floated, transferred to draining molds, cooled, and ultimately packaged (Mangione et al., 2023).

The significance of ricotta whey cheese extends beyond its culinary appeal. Rich in nutritional value, this product boasts a noteworthy content of whey proteins, particularly those abundant in sulfur-

* Corresponding author.
E-mail address: alessandra.biancolillo@univaq.it (A. Biancolillo).

https://doi.org/10.1016/j.jfca.2024.107019
Received 22 September 2024; Received in revised form 28 October 2024; Accepted 23 November 2024
Available online 26 November 2024

containing amino acids, coupled with a relatively low-fat composition. Its versatility allows for variations such as ripened or smoked ricotta, yet it is commonly marketed as a fresh product.

One of the distinguishing features of ricotta whey cheese lies in its adaptability to various mammalian milk sources, each imparting distinct nutritional and organoleptic characteristics to the final product. This flexibility in milk selection adds layers of complexity to the cheese's flavor profile and nutritional composition, making ricotta a canvas for diverse sensory experiences.

In Italy, the authenticity and origin of ricotta whey cheese, especially the "Ricotta Romana," derived exclusively from 100 % sheep's whey, take on a particular significance. Recognized as a Protected Designation of Origin (PDO) product, sheep ricotta is produced in the Lazio Region (Central Italy) and it carries a cultural and historical legacy that demands preservation and authentication. The PDO designation ensures that this whey cheese is produced in a specific geographic area using traditional methods, underscoring the importance of maintaining the integrity of regional culinary heritage.

Dairy products in general, and PDO cheeses in particular, hold heightened commercial value, making them susceptible to fraudulent practices, including mislabeling the animal/geographical origin of whey. While it is established that the type of whey significantly influences the sensory characteristics of ricotta whey cheeses (Miele et al., 2021), various further aspects, such as local processing techniques (Aït-Kaddour et al., 2021; Mandal et al., 2023; Ortiz Araque et al., 2018), the breed and diet of dairy animals (Fusaro et al., 2019; Muniz de Souza et al., 2021; Zeppa et al., 2005), and the general composition (da Silva Medeiros et al., 2024; da Silva Medeiros et al., 2023; Karaziack et al., 2024; Manuelian et al., 2017) donate to the unique origin-related traits of PDO-designated products.

In the case of PDO Ricotta Romana, the whey cheese is crafted from full-fat sheep whey sourced from the prevalent breeds in the production area. Its peculiar fairly sweet flavor is a result of the specific feed provided to the milk ewes, consisting of forage from natural pastures, meadows, and the characteristic grasslands within Lazio. This underscores that authenticating a PDO ricotta whey cheese extends beyond identifying the whey origin; discrimination between PDO and non-PDO products, even when utilizing the same type of whey, becomes imperative.

Existing discussions highlight various methods proposed to detect the undisclosed addition of cow whey to whey cheeses originating from different animals (Fagnani et al., 2022). Additionally, instrumental methods coupled with chemometric approaches have been employed to address geographical traceability and authentication in numerous dairy products (Kamal and Karoui, 2015). However, to the best of our knowledge, analytical methods specific to the geographical traceability of ricotta whey cheeses, particularly PDO specialties, and strategies for discriminating ricotta whey cheeses based on whey origin remain not widely discussed in the literature.

A previous work (Biancolillo et al., 2022b) has indicated how fatty acid profiles can be used for the classification of ricotta cheeses according to the animal origin and the PDO label. Nevertheless, in the current research study, the possibility of achieving this aim by exploiting a simple, fast and non-destructive (or semi-destructive) technique such as Attenuated Total Reflection-Fourier-transform infrared (FTIR) spectroscopy ATR-FTIR has been taken into consideration. To achieve this goal, this study integrates FT-IR with a linear discriminant classifier (Sequential Preprocessing through ORThogonalization Linear Discriminant Analysis, SPORT-LDA) and a class-modelling approach (Soft and Independent modeling of Class Analogies, SIMCA), leveraging the proven efficiency of these strategies in similar contexts (Biancolillo et al., 2022a; Di Donato et al., 2022; Foschi et al., 2023; Mellado-Carretero et al., 2020; Petrakis and Polissiou, 2017; Song et al., 2020).

2. Materials and methods

2.1. Samples

A total of 287 samples of ricotta cheese were available for analysis. Of these, 191 were made from sheep milk, 41 from cow milk, 31 from goat milk, and 24 from water buffalo milk.

2.2. ATR-FTIR analysis

All the available samples were analyzed using a PerkinElmer Spectrum Two™ FT-IR spectrometer, featured with deuterated triglycine sulfate (DTGS) detector, a PerkinElmer Universal Attenuated Total Reflectance (uATR) sampling accessory, and an integrated load monitoring system for applying controlled pressure to the sample, ensuring its adhesion to the diamond. The spectrometer recorded spectra within the range of 4000–500 cm^{-1} at a resolution of 4 cm^{-1}, following an air background measurement at ambient temperature. Each cheese sample was subjected to two replicate analyses (averaged before the chemometric analysis). MIR spectra were exported to MATLAB 2015b and transformed into pseudo-absorbance (log(1/R)). All the data analysis was run using in-house functions.

2.3. Chemometric classifiers, model calibration and validation

As displayed in the schematic workflow shown in Fig. 1, two different classification problems have been faced. In the first case, Sequential Preprocessing through ORThogonalization Linear Discriminant Analysis (SPORT-LDA) (Roger et al., 2020) was used to discriminate all the samples according to the nature of the milk they were made of. This method is a discriminant approach derived from SOPLS-LDA (Biancolillo et al., 2015) and utilizes the concept of sequential multi-block modeling to integrate various data preprocessing techniques. Assuming two data blocks $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$, used to predict a response $\boldsymbol{Y}$, the SO-PLS algorithm can be summarized in the following four steps. At first $\boldsymbol{X_1}$ is used to predicted $\boldsymbol{Y}$ by Partial Least Squares (PLS). Then, $\boldsymbol{X_2}$ is orthogonalized with respect to the $\boldsymbol{X_1}$-scores. The orthogonalized block is then used to predict the residuals obtained in the first regression. Eventually, the response is globally predicted by summing up all the contributions. Classification is then achieved by applying LDA on the predicted $\boldsymbol{Y}$. In a case where more than two data blocks are involved, each block is orthogonalized with respect to all the previously modelled predictors and used to predict the residuals obtained at the previous regression step. A more detailed description of the algorithm can be found here. SPORT-LDA exploits the same algorithm; the main difference between these approaches is that SO-PLS handles data blocks of different natures as input blocks (Biancolillo and Næs, 2019a). In SPORT the same data block is pretreated using diverse preprocessing approaches; the resulting data matrices (which will be as many as the number of pretreatments used) will be used modelled.

The tested pretreatments were Mean-Centering (MC), Standard Normal Variate (SNV) (Barnes et al., 1989), and the first (D1) and second (D2) derivatives (Savitzky and Golay, 1964). SPORT-LDA was applied to extract information from four distinct data blocks: $\boldsymbol{X}_1$ is the matrix pretreated by MC, $\boldsymbol{X}_2$ is the one preprocessed by SNV+MC, $\boldsymbol{X}_3$ is pretreated by D1+MC, and $\boldsymbol{X}_4$ is preprocessed by D2+MC. All possible combinations of latent variables (LVs) within the range of 1–10 were tested. Afterward, SOPLS is applied to construct a regression model that correlates these data blocks with a $\boldsymbol{Y}$ dummy matrix encoding the class information. Subsequently, LDA is performed on the $\boldsymbol{Y}$ values predicted by the regression model. Once the final calibration model is established, the approach can be used to predict unknown/test samples.

Linear Discriminant Analysis (LDA) is a well-known classification method that has been widely applied since its introduction by Fisher in 1936 (FISHER, 1936). The technique assumes that the samples from each class are distributed normally and share a common

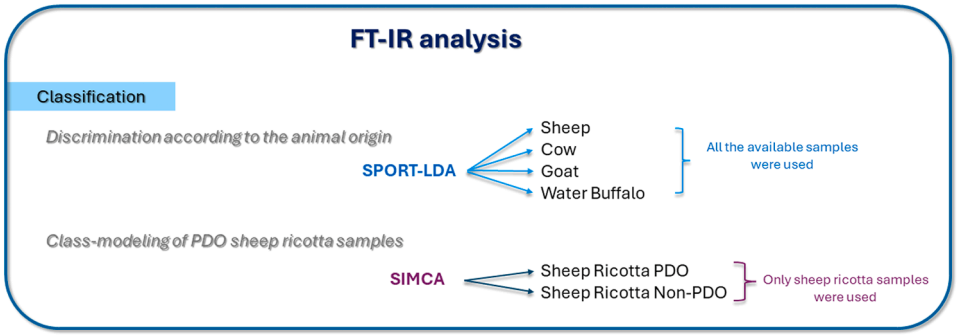

Fig. 1. Schematic workflow of the study.

variance-covariance matrix. Based on these assumptions, LDA seeks to model the likelihood that a sample belongs to a given class by calculating the probability associated with it. This is done by determining the distance between the sample and the centroid of the class in multivariate space. The sample is then assigned to the class for which this probability is highest. The application of LDA leads to the estimation of the Canonical VAriates (CVAs); i.e., the directions that effectively enhance group separability in the new transformed space. This is achieved by maximizing the ratio of between-group variance (i.e., the variance among different classes) to within-group variance (the intra-class variance). The reader is addressed to the literature for more details on LDA.

For further details on the algorithm, readers are encouraged to consult the relevant literature (Biancolillo and Næs, 2019b; Næs et al., 2011). In the second part of the study, the analysis focuses on samples made from sheep milk. A class-modeling approach was employed to classify PDO cheese against samples without this quality mark.

The method used for this classification is SIMCA (Wold, 1976; Wold and Sjöström, 1977), which relies on the creation of a PCA model on samples belonging to the class of interest. Once the class-space is defined, the distance (d_i) of each analyzed sample within this PCA-defined space is calculated. This distance considers both the principal component space (The squared Mahalanobis distance that follows the T^2 statistics) and the residuals from the PCA model (The squared orthogonal Euclidean distance between the sample and its projection into the PCA model, which follows the Q statistics). d_i is calculated as reported in Eq. 1:

$$d_i = \sqrt{T^2_{i,red} + Q^2_{i,red}} \leq \sqrt{2} \quad (1)$$

In this equation, $T^2_{i,red}$ represents the ratio of T^2_i to T^2_{crit} and $Q_{i,red}$ denotes the ratio of Q_i and Q_{crit}, where T^2_{crit} and Q_{crit} are the critical values corresponding to a 95 % confidence level.

Thus, if a sample's distance exceeds a certain threshold (often set at the square root of 2 for specific reasons (Yue and Qin, 2001)), it is rejected by the class model and predicted as not belonging to that class. Conversely, if the distance is below the cut-off, the object is accepted and considered to appertain to the modeled category.

Regardless of the classifier used, the calibration model parameters were determined through internal validation (7-fold cross-validation), while the predictive capabilities of the models were assessed through external validation. Therefore, before modeling, the data were split into a training (or calibration) and a test (or validation) set using the Duplex algorithm (Snee, 1977). In particular, the training set was made of 131 sheep milk samples, 26 cow milk samples, 21 goat milk samples, and 12 water buffalo milk samples, for a total of 190 calibration objects, while the test set was constituted of 60 sheep milk samples, 15 cow milk samples, 10 goat milk samples, and 12 water buffalo milk samples for a total of 97 validation objects. For optimizing the model parameters in SPORT-LDA, cross-validation was employed to determine the optimal number of latent variables (LVs) to extract from each block. Several models were created, testing all possible combinations of LVs across the different blocks, with a maximum of 10 LVs extracted from each of them.

The chosen calibration model was the one that achieved the best classification rate in cross-validation. Descriptions and examples of how to perform this procedure can be found in (Biancolillo and Næs, 2019b).

For SIMCA, the optimization of parameters, i.e., the number of principal components (PCs) and the optimal data pretreatment, involved the creation of several SIMCA models using differently preprocessed data. These models were calculated by extracting an increasing number of PCs, ranging from 1 to 10. The optimal model was defined based on its efficiency, measured as the geometric mean of sensitivity (the percentage of samples correctly accepted by the class-model) and specificity (the percentage of samples correctly rejected by the class-model).

3. Results

In Fig. 2, the average spectra associated with the relevant categories are shown. In particular, in Fig. 2A, the mean spectrum associated with the ricotta cheeses produced with the different milks is shown. On the other hand, Fig. 2B depicts the average spectrum for PDO and non-PDO samples, regarding the nature of the milk.

From the figure (both A and B) it is clear that the spectra present similar profiles. This is given by the fact that the main constituents of the ricotta are the same, both when looking at the spectra of the different animals and between PDO and non-PDO samples.

The analysis of ricotta samples using spectroscopic techniques provides valuable insights into their molecular composition. Although the spectra do not reveal clear differences among the three varieties of ricotta, the identification and characterization of specific vibrational modes are crucial. The spectral regions of interest, particularly between 3500 and 2800 cm^{-1} and 1750–1000 cm^{-1}, highlight the presence of various functional groups and their respective vibrational modes. The OH and NH stretching vibrations observed in the 3325–3081 cm^{-1} range indicate the presence of water and proteins, essential components of ricotta. The peaks at 2918 and 2851 cm^{-1} correspond to the stretching vibrations of CH_2 and CH_3 groups, reflecting the fatty acid content. The significance of the peak at 1741 cm^{-1} lies in its association with the ester C═O bond of lipids, an essential aspect of the cheese's fat content. Furthermore, the region between 1700 cm^{-1} and 1500 cm^{-1} is critical as it includes the characteristic peaks of amide groups in peptides, sensitive to the conformation assumed by the protein backbone. The peak at 1642 cm^{-1} (amide I) is related to the amide C═O bond, while the 1578 cm^{-1} peak (amide II) corresponds to the mixed vibration of N-H bending and C-N stretching in the peptide bond. In the lower absorption region (1500–1100 cm^{-1}), the observed signals provide information about the carbohydrate content and the ester bonds in triglycerides, contributing to the overall nutritional profile of the ricotta. This detailed spectral analysis is instrumental in comprehensively understanding the molecular composition and potential quality differences in ricotta varieties, even if the differences are subtle and require further

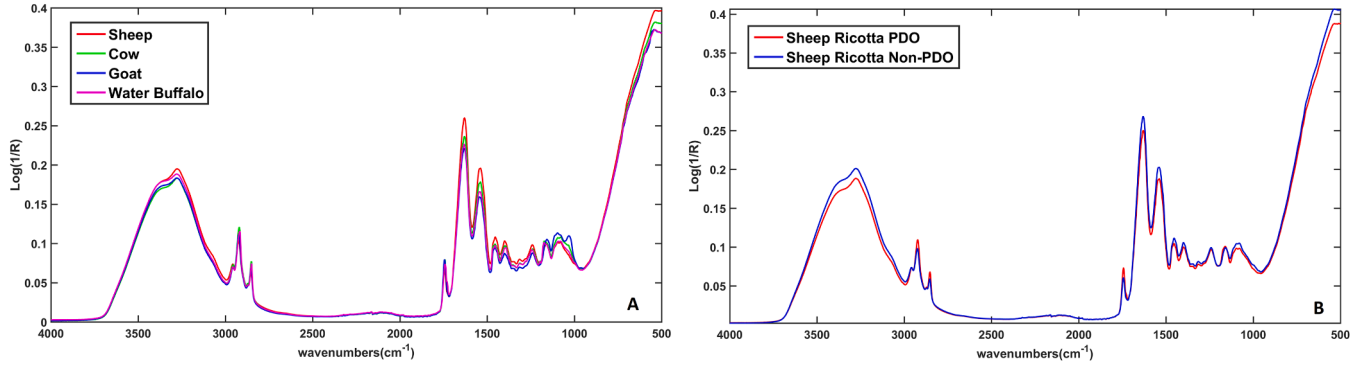

Fig. 2. Raw ATR-FTIR Spectra. A) Mean spectra of sheep (red solid line), cow (green solid line), goat (blue solid line), and water buffalo (magenta solid line); B) sheep ricotta PDO (red solid line), and sheep ricotta Non-PDO (blue solid line).

investigation to be distinguished (Chen and Irudayaraj, 1998; Jesus et al., 2020).

3.1. Classification of Ricotta cheeses according to the animal origin of the milk

As previously mentioned, initially, all available samples were classified using SPORT-LDA based on the different types of milk used in the ricotta preparation. As outlined in Section 2.3, the samples were divided into a training set and a test set.

The model that achieved the highest accuracy in cross-validation utilized 5, 3, 6, and 3 LVs from $\boldsymbol{X}_1$(MC), $\boldsymbol{X}_2$(SNV + MC), $\boldsymbol{X}_3$(D1+MC), and $\boldsymbol{X}_4$(D2+MC), respectively, resulting in an accuracy of 98 %. When this model was applied to the test set, it correctly classified 97 % of the samples, as detailed in the confusion matrix shown in Table 1. From this is evident that, over 97 validation samples, only 3 objects have been misclassified (one belonging to Class Sheep, one appertaining to Class Cow, and one from Class Water Buffalo), being all of them erroneously assigned to Class Goat.

Fig. 3 has been realized in order to appreciate the differences among the diverse classes when training (empty symbols) and test (filled symbols) samples are projected onto the space spanned by canonical variates (CVs). In particular, in Fig. 3A the objects are displayed in the space defined by the three CVs. To facilitate visualization, in subplots B, C, and D, the projection into the CV1-CV2, CV1-CV3, and CV2-CV3 spaces is also shown. In Fig. 3, the legend is the same in all subplots.

The inspection of Fig. 3B reveals that the main discrimination that takes place along CV1 is the one between cow (green diamonds, at negative CV1 values) and water buffalo (magenta stars, at positive CV1 values) samples. This latter category is well-discerned by all the others along CV2. From Fig. 3C it emerges that a slight discrimination between Sheep ricotta and water buffalo samples with respect to the other two classes can be observed on CV1 when individuals are projected in the space CV1-CV3. The third component allows the discrimination of Goat ricotta samples (blue downward triangles, at positive CV3 values) with respect to all the others. Fig. 3D mainly enhances the peculiarity of water buffalo samples (at negative CV2 values) with respect to all the other categories that present slightly negative or positive scores on this component. In this case, CV3 does not provide additional information about the different clusters.

SPORT-LDA allows for the use of VIP analysis in its embedded version as suggested by Biancolillo et al. (Biancolillo et al., 2016). The VIP analysis is a valuable tool that provides a measure of the contribution of each variable to the model, allowing the detection of the most influential variables in a predictive model. Customarily, these are identified as those with scores greater than 1.

This approach has revealed that, regardless of the animal source of the milk, the same spectral variables consistently emerged as relevant. The VIP analysis highlighted three critical spectral ranges: 3300–3100 cm^{-1}, 2900–2800 cm^{-1}, and 1700–1300 cm^{-1}.

The significance of the first reported spectral region (3300–3100 cm^{-1}) is expected, given that ricotta cheese is rich in proteins like albumin and globulin. The importance of this range across different animal sources suggests that protein content and hydration levels are crucial factors in the classification of ricotta. While the fatty acid composition can vary between ricotta made from different types of milk (Biancolillo et al., 2022b), the overall lipid content remains a key distinguishing feature. The high VIP scores in the range of 2900–2800 cm^{-1} emphasize the role of fat content in differentiating ricotta samples, highlighting its relevance regardless of the milk source. Eventually, the broad spectral range of 1700–1300 cm^{-1}, indicative of protein secondary structures and various organic compounds within the cheese, points to the complex interplay of protein structures and other organic components that define the unique characteristics of ricotta cheese.

The consistent identification of these spectral variables as significant, irrespective of the animal source, underscores the effectiveness of the SPORT-LDA model in capturing the essential compositional features of ricotta whey cheese. These findings validate the use of ATR-FTIR spectroscopy and SPORT-LDA for cheese classification and provide a deeper understanding of the molecular characteristics that distinguish ricotta from different milk sources.

Table 1
SPORT-LDA: Data Fusion matrix obtained predicting the external test set.

	Pred Class Sheep	Pred Class Cow	Pred Class Goat	Pred Class Water Buffalo
Class Sheep	59	0	1	0
Class Cow	0	14	1	0
Class Goat	0	0	10	0
Class Water Buffalo	0	0	1	11

3.2. Classification of PDO and Non-PDO sheep ricotta cheeses

In the second part of the study, the focus narrowed to sheep's ricotta samples to determine if it is possible to differentiate between PDO and non-PDO samples. Therefore, the 191 sheep ricotta samples were taken into account. The spectra associated with these observations were divided into a training and a test set of 134 and 57 samples, respectively. The calibration set contained 67 PDO and 67 Non-PDO individuals, while the test was constituted by 29 PDO and 28 Non-PDO.

Given the asymmetric nature of the classification problem, SIMCA was employed to address it. Various pretreatments were evaluated to identify the most suitable one, including MC, SNV(+MC), D1(+MC), D2 (+MC), SNV+D1(+MC), and SNV+D2(+MC). This approach resulted in six different SIMCA models, with the details provided in Table 2.

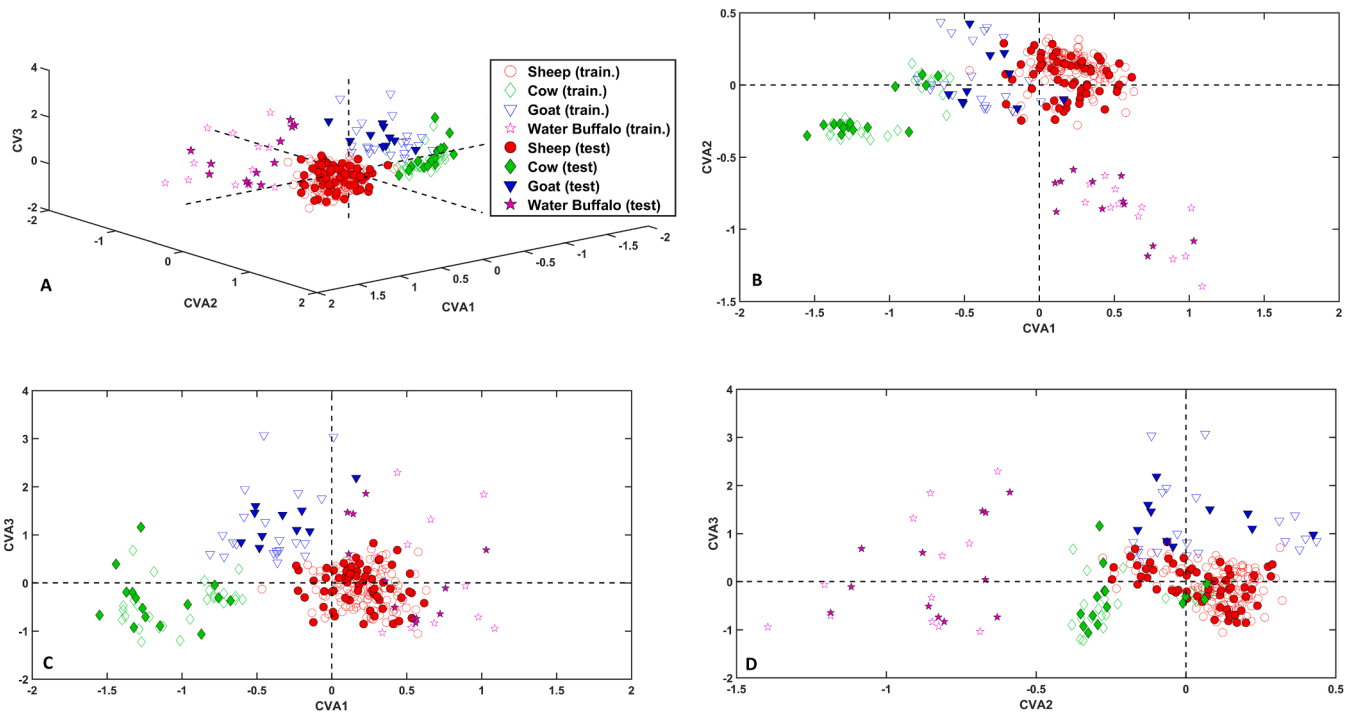

Fig. 3. SPORT-LDA: graphical representation of the results. A) Projection of samples in the space spanned by CV1, CV2, and CV3; B) Projection of samples in the space spanned by CV1 and CV2; C) Projection of samples in the space spanned by CV1 and CV3; D) Projection of samples in the space spanned by CV2 and CV3. Legend: Red circles: sheep ricotta; green diamonds: cow ricotta; blue downward triangles: goat ricotta; magenta stars: water buffalo ricotta.

Table 2
SIMCA: Data pretreatment, number of extracted PCs, and cross-validated efficiencies (%CV).

Preprocessing	PCs	Efficiency (%CV)
Mean Centering (MC)	10	80.6
SNV (+MC)	8	80.0
D1 (+MC)	9	79.0
D2 (+MC)	11	77.3
SNV + D1 (+MC)	10	80.3
SNV + D2 (+MC)	10	78.4

The calibration model chosen as optimal was the one based on data pre-treated with SNV. This decision has been taken because it provided one of the highest efficiencies (80 %) exploiting the most parsimonious number of PCs. Applying this model to the test samples achieved a sensitivity of 82.1 % and a specificity of 82.7 %.

The graphical representation of the SIMCA outcome is shown in Fig. 4. The plot illustrates the projection of the samples, sheep ricotta PDO (indicated by red stars) and sheep ricotta Non-PDO (indicated by black downward triangles), onto the T^2_{red} and Q_{red} space. The black dotted line represents the critical distance at the 95 % confidence level. This plot clearly shows that the majority of the PDO samples, located

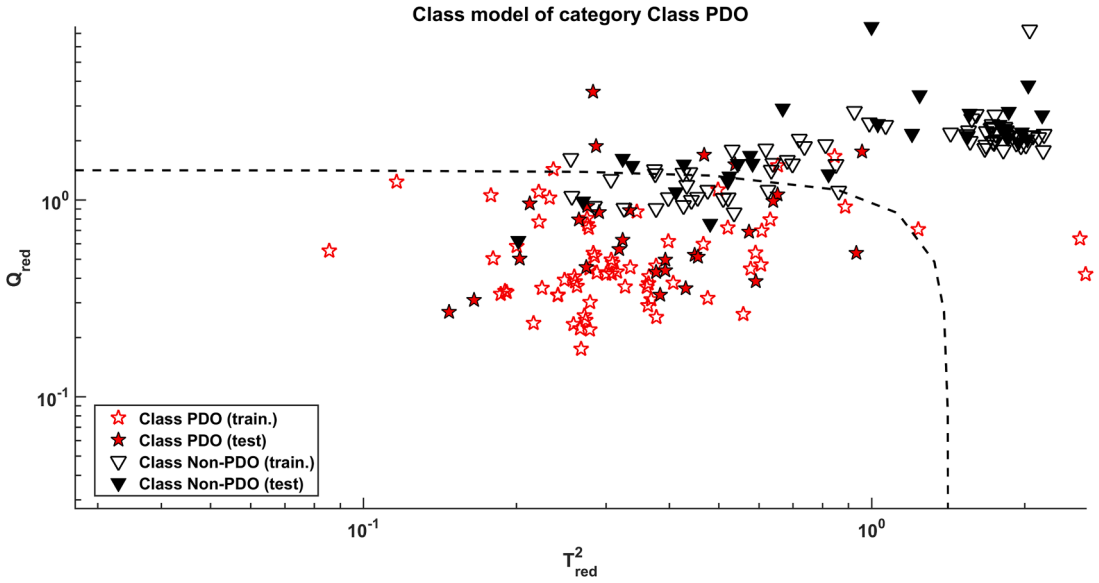

Fig. 4. SIMCA analysis: T^2_{red} vs Q_{red} plot. Legend: Red stars: sheep ricotta PDO; Black downward triangles: sheep ricotta Non-PDO. Empty symbols: training samples; Filled symbols: test samples. The dashed line represents the acceptance/rejection threshold.

within the critical distance, have been correctly classified by the model. Specifically, only 5 PDO samples have been erroneously rejected, while 5 Non-PDO samples have been mistakenly accepted.

Eventually, in order to test the applicability of the model, a more rigorous approach has been applied and calculations re-run. In particular, samples have been re-split decreasing the difference between the number of calibration and validation samples. 100 individuals were used to train the SIMCA model, and 91 to test it. This led to cross-validated efficiencies that span between 80.8 % (SNV+MC, 9PCs) and 68.6 % (SNV+D2+MC, 11 PCs), confirming the use of SNV as data pretreatment. The application of this model to the test set led to 82.2 % of sensitivity and 84.8 % of specificity, indicating the suitability of the approach for this purpose.

4. Conclusions

The classification of ricotta whey cheese from different milk sources and PDO statuses through ATR-FTIR and chemometrics has provided significant insights into the compositional variations and potential quality markers of this traditional dairy product. Our results demonstrate the capability of FT-IR spectroscopy combined with chemometric techniques to distinguish between ricotta samples based on their animal origin and/or production methods. This ability to classify ricotta samples accurately is critical for maintaining the integrity of PDO products and ensuring consumer trust in these labels.

Several key findings emerged from the present study. First, the spectral data revealed distinct differences in the protein and fat content of ricotta cheeses derived from different milk sources. For instance, ricotta made from sheep's milk exhibited higher levels of specific fatty acids compared to those made from cow's milk. These differences are not only important for nutritional profiling but also play a crucial role in the sensory attributes of the cheese, such as texture and flavor.

Moreover, the chemometric analysis underscored the effectiveness of this approach in managing the complexity of the spectral data. Both SPORT-LDA and SIMCA were particularly useful in identifying patterns and correlations within the dataset, allowing us to classify the ricotta samples with high accuracy. The robustness of these techniques suggests their potential application in quality control processes within the dairy industry.

It is also noteworthy that the PDO status significantly influenced the compositional characteristics of the ricotta samples. PDO ricotta exhibited a more uniform composition, likely reflecting the stringent production standards required for this designation. This uniformity is a testament to the effectiveness of PDO regulations in ensuring product consistency and quality.

However, some limitations should be acknowledged. The study primarily focused on the chemical composition of the ricotta samples and further research is needed to explore other factors that may influence classification, such as microbial content and environmental conditions during production. Additionally, expanding the sample size and diversity could enhance the generalizability of our findings.

In conclusion, our study demonstrates that ATR-FTIR spectroscopy combined with chemometric techniques is a powerful tool for the classification and quality assessment of ricotta whey cheese. The ability to accurately differentiate between ricotta samples from various milk sources and with different PDO statuses highlights the potential of these methods in supporting the authenticity and quality control of dairy products.

Our findings underline the importance of rigorous analytical techniques in maintaining the integrity of traditional food products. As consumer demand for authentic and high-quality foods continues to grow, the adoption of advanced analytical methods will be essential for producers and regulatory bodies alike.

Future research should aim to address the identified limitations by incorporating a broader range of samples and exploring additional variables that may impact ricotta classification. By doing so, we can further refine these techniques and enhance their applicability in the dairy industry.

Overall, this study contributes to a better understanding of the compositional diversity of ricotta whey cheese and supports the ongoing efforts to protect and promote traditional food products through scientific innovation.

CRediT authorship contribution statement

Samantha Reale: Supervision, Funding acquisition. **Francesco Poles:** Investigation, Formal analysis. **Martina Foschi:** Writing – original draft, Visualization, Validation, Supervision, Software, Formal analysis, Data curation. **Alessandra Biancolillo:** Writing – original draft, Visualization, Validation, Supervision, Software, Investigation. **Angelo Antonio D'Archivio:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Aït-Kaddour, A., Hassoun, A., Bord, C., Schmidt-Filgueras, R., Biancolillo, A., Di Donato, F., Temiz, H.T., Cozzolino, D., 2021. Application of spectroscopic techniques to evaluate heat treatments in milk and dairy products: an overview of the last decade. Food Bioprocess Technol. 14, 781–803. https://doi.org/10.1007/s11947-021-02607-0.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43, 772–777. https://doi.org/10.1366/0003702894202201.

Biancolillo, A., Foschi, M., Di Micco, M., Di Donato, F., D'Archivio, A.A., 2022a. ATR-FTIR-based rapid solution for the discrimination of lentils from different origins, with a special focus on PGI and Slow Food typical varieties. Microchem. J. 178. https://doi.org/10.1016/j.microc.2022.107327.

Biancolillo, A., Liland, K.H., Måge, I., Næs, T., Bro, R., 2016. Variable selection in multi-block regression. Chemom. Intell. Lab. Syst. 156, 89–101. https://doi.org/10.1016/j.chemolab.2016.05.016.

Biancolillo, A., Måge, I., Næs, T., 2015. Combining SO-PLS and linear discriminant analysis for multi-block classification. Chemom. Intell. Lab. Syst. 141, 58–67. https://doi.org/10.1016/j.chemolab.2014.12.001.

Biancolillo, A., Næs, T., 2019a. The sequential and orthogonalized pls regression for multiblock regression: theory, examples, and extensions. Data Handl. Sci. Technol. https://doi.org/10.1016/B978-0-444-63984-4.00006-5.

Biancolillo, A., Næs, T., 2019b. The sequential and orthogonalized PLS regression for multiblock regression: theory, examples, and extensions. Data Handl. Sci. Technol. https://doi.org/10.1016/B978-0-444-63984-4.00006-5.

Biancolillo, A., Reale, S., Foschi, M., Bertini, E., Antonelli, L., D'Archivio, A.A., 2022b. Characterization and authentication of "ricotta" whey cheeses through GC-FID analysis of fatty acid profile and chemometrics. Molecules 27. https://doi.org/10.3390/molecules27217401.

Chen, M., Irudayaraj, J., 1998. Sampling technique for cheese analysis by FTIR spectroscopy. J. Food Sci. 63, 96–99. https://doi.org/10.1111/j.1365-2621.1998.tb15684.x.

da Silva Medeiros, M.L., Freitas Lima, A., Correia Gonçalves, M., Teixeira Godoy, H., Fernandes Barbin, D., 2023. Portable near-infrared (NIR) spectrometer and chemometrics for rapid identification of butter cheese adulteration. Food Chem. 425, 136461. https://doi.org/10.1016/j.foodchem.2023.136461.

da Silva Medeiros, M.L., Moreira de Carvalho, L., Madruga, M.S., Rodríguez-Pulido, F.J., Heredia, F.J., Fernandes Barbin, D., 2024. Comparison of hyperspectral imaging and spectrometers for prediction of cheeses composition. Food Res. Int. 183, 114242. https://doi.org/10.1016/j.foodres.2024.114242.

Di Donato, F., Biancolillo, A., Foschi, M., D'Archivio, A.A., 2022. Application of SPORT algorithm on ATR-FTIR data: A rapid and green tool for the characterization and discrimination of three typical Italian Pecorino cheeses. J. Food Compos. Anal. 114. https://doi.org/10.1016/j.jfca.2022.104784.

Fagnani, R., Damião, B.C.M., Trentin, R.P.S., Vanot, R.L., 2022. Predicting adulteration of grated Parmigiano Reggiano cheese with Ricotta using electrophoresis, multivariate nonlinear regression and computational intelligence methods. Int. J. Dairy Technol. 75, 239–245. https://doi.org/10.1111/1471-0307.12818.

FISHER, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7, 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Foschi, M., Tozzi, L., Di Donato, F., Biancolillo, A., D'Archivio, A.A., 2023. A novel FTIR-based chemometric solution for the assessment of saffron adulteration with non-fresh stigmas. Molecules 28. https://doi.org/10.3390/molecules28010033.

Fusaro, I., Giammarco, M., Chincarini, M., Odintsov Vaintrub, M., Formigoni, A., Mammi, L.M.E., Vignola, G., 2019. Fatty acids, health indices and sensory properties of Ricotta cheese from sheep fed three different diets. Int. J. Dairy Technol. 72, 427–434. https://doi.org/10.1111/1471-0307.12613.

Jesus, J.C., de, Silva, G.J., Gonçalves, B.-H.R.F., Souza, M.R., de, Santos, L.S., Ferrão, S.P. B., 2020. Quick identification of the time of maturation of artisanal minas cheese by FTIR-ATR spectroscopy and multivariate techniques. J. Braz. Chem. Soc. 31, 2000–2011. https://doi.org/10.21577/0103-5053.20200100.

Kamal, M., Karoui, R., 2015. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: a review. Trends Food Sci. Technol. 46, 27–48. https://doi.org/10.1016/j.tifs.2015.07.007.

Karaziack, C.B., Vidal, C., Pasquini, C., Barbin, D.F., Viotto, W.H., 2024. Application of near-infrared hyperspectral imaging for determination of cheese chemical composition. J. Food Compos. Anal. 127, 105994. https://doi.org/10.1016/j.jfca.2024.105994.

Mandal, R., Karmakar, P., Roy, K., Rani, R., 2023. Heat and acid coagulated milk products: Physicochemical changes during processing and storage. Chem. Milk. Milk. Prod.: Physicochem. Prop., Ther. Charact., Process. Methods.

Mangione, G., Caccamo, M., Natalello, A., Licitra, G., 2023. Graduate Student Literature Review: History, technologies of production, and characteristics of ricotta cheese. J. Dairy Sci. 106, 3807–3826. https://doi.org/10.3168/jds.2022-22460.

Manuelian, C.L., Currò, S., Visentin, G., Penasa, M., Cassandro, M., Dellea, C., Bernardi, M., De Marchi, M., 2017. Technical note: At-line prediction of mineral composition of fresh cheeses using near-infrared technologies. J. Dairy Sci. 100, 6084–6089. https://doi.org/10.3168/jds.2017-12634.

Mellado-Carretero, J., García-Gutiérrez, N., Ferrando, M., Güell, C., García-Gonzalo, D., De Lamo-Castellví, S., 2020. Rapid discrimination and classification of edible insect powders using ATR-FTIR spectroscopy combined with multivariate analysis. J. Insects Food Feed 6, 141–148. https://doi.org/10.3920/jiff2019.0032.

Miele, N.A., Puleo, S., Di Monaco, R., Cavella, S., Masi, P., 2021. Sensory profile of protected designation of origin water buffalo ricotta cheese by different sensory methodologies. J. Sens. Stud. 36, e12648. https://doi.org/10.1111/joss.12648.

Muniz de Souza, M.Y., Cavalcanti, F.B., dos Santos Pereira, E.V., Alonso Buriti, F.C., Florentino, E.R., 2021. Ricotta cream: classification based on moisture and fat content considering general standards for cheeses and cream cheeses. Heliyon 7, e08408. https://doi.org/10.1016/j.heliyon.2021.e08408.

Næs, T., Måge, I., Segtnan, V.H., 2011. Incorporating interactions in multi-block sequential and orthogonalised partial least squares regression. J. Chemom. 25, 601–609. https://doi.org/10.1002/cem.1406.

Ortiz Araque, L.C., Darré, M., Ortiz, C.M., Massolo, J.F., Vicente, A.R., 2018. Quality and yield of Ricotta cheese as affected by milk fat content and coagulant type. Int. J. Dairy Technol. 71, 340–346. https://doi.org/10.1111/1471-0307.12431.

Petrakis, E.A., Polissiou, M.G., 2017. Assessing saffron (Crocus sativus L.) adulteration with plant-derived adulterants by diffuse reflectance infrared Fourier transform spectroscopy coupled with chemometrics. Talanta 162, 558–566. https://doi.org/10.1016/j.talanta.2016.10.072.

Roger, J.-M., Biancolillo, A., Marini, F., 2020. Sequential preprocessing through ORThogonalization (SPORT) and its application to near infrared spectroscopy. Chemom. Intell. Lab. Syst. 199. https://doi.org/10.1016/j.chemolab.2020.103975.

Savitzky, Abraham, Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047.

Snee, R.D., 1977. Validation of Regression Models: Methods and Examples. Technometrics 19, 415–428. https://doi.org/10.1080/00401706.1977.10489581.

Song, W., Jiang, N., Wang, H., Guo, G., 2020. Evaluation of machine learning methods for organic apple authentication based on diffraction grating and image processing. J. Food Compos. Anal. 88. https://doi.org/10.1016/j.jfca.2020.103437.

Wold, S., 1976. Pattern recognition by means of disjoint principal components models. Pattern Recognit. 8, 127–139. https://doi.org/10.1016/0031-3203(76)90014-5.

Wold, S., Sjöström, M., 1977. SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy. In: Chemometrics: Theory and Application, ACS Symposium Series. AMERICAN CHEMICAL SOCIETY, pp. 12–243. https://doi.org/10.1021/bk-1977-0052.ch012.

Yue, H.H., Qin, S.J., 2001. Reconstruction-based fault identification using a combined index. Ind. Eng. Chem. Res. 40, 4403–4414. https://doi.org/10.1021/ie000141+.

Zeppa, G., Giordano, M., Bertolino, M., Gerbi, V., 2005. Application of artificial neural network on mono- and sesquiterpenes compounds determined by headspace solid-phase microextraction-gas chromatography-mass spectrometry for the Piedmont ricotta cheese traceability. J. Chromatogr. A 1071, 247–253. https://doi.org/10.1016/j.chroma.2004.11.083.