https://doi.org/10.1016/j.jfca.2023.105211

Application of NIR spectroscopy coupled with DD-SIMCA class modelling for the authentication of pork meat

Michela Pia Totaro, Giacomo Squeo*, Davide De Angelis, Antonella Pasqualone, Francesco Caponio, Carmine Summo

University of Bari "Aldo Moro", Dept. of Soil, Plant and Food Sciences (Di.S.S.P.A.), via Amendola 165/A, 70126 Bari, Italy

*Corresponding author: giacomo.squeo@uniba.it, Department of Soil Plant and Food Sciences, University of Bari Aldo Moro, Via Amendola 165/A, 70126 Bari, Italy

Email addresses:
michela.totaro@uniba.it; giacomo.squeo@uniba.it; davide.deangelis@uniba.it; antonella.pasqualone@uniba.it; francesco.caponio@uniba.it; carmine.summo@uniba.it

Abstract

Rearing system, breed, geographical origin, and animal welfare are some of the more recent quality features able to drive consumer's choice in meat and meat products purchase. The same aspects are marketing leverage from the producer's side. However, there is a need to develop rapid analytical strategies to check the compliance of declared quality with actual quality features. To face with this issue, this work aims at test the feasibility of using the fat portion as a marker of authenticity. To this end, fat samples from extensively and intensively reared pigs have been collected and analysed by NIRS (Near Infra-Red Spectroscopy). Then, after multivariate exploration, the Data Driven variant of Soft Independent Modelling of Class Analogy (DD-SIMCA) was used to develop models for the target class. The type of signal pre-processing and the number of principal components (PCs) used strongly affected the performance of the classification. Excellent results were obtained with SNV pre-treated data and 4 PCs that allowed to reach 100% sensitivity and specificity in calibration and validation.

Keywords: Chemometrics; vibrational spectroscopy; pig; fat; food analysis; food composition; PDO; rearing system.

1. Introduction

One of the most commonly consumed food in Western countries is meat, since it is an excellent source of protein and essential nutrients (Mabood et al., 2020). When consumers purchase meat, factors such as health implications, price, quality, origin, and authenticity are important decision drivers (Prieto et al., 2017).

Nowadays, consumers pay much more attention to the quality aspects related to the intrinsic characteristics of the animals and of the rearing system, with a preference for autochthonous species and breeds, geographical origin and feeding systems (Prieto et al., 2017). At the same time, food producers try to maximise the market potential of their products, leveraging for example, on its traditional nature or the sustainability of the production system or the nutritional profile of their products (Monahan et al., 2018) and the animal welfare, such as absence of thirst, hunger, discomfort, disease, pain and injuries, stress and the expression of normal behaviour (Farm Animal Welfare Council, 1992; Council Regulation (EC) No. 1099/2009).

However, the compliance of a product with the declared or supposed quality features is mostly of a documentary nature, which results in low ranking on the valorisation chain and in possible consumer distrust. As proof of this, according to the last annual report from the EU Food Fraud Network, meat and meat products are the third most reported category concerning non-compliance notifications. These range from incomplete labels concerning ingredients and traceability, to incomplete or missing documentation (European Union, 2022). As for suspicious fraud notifications, meat and meat products rank third again and, as a matter of concern, it should be noted that the most frequent non-compliances related to fraud suspicions are "absent/falsified/manipulated documentation" (European Union, 2022).

Bearing this framework in mind, both authorities and consumers could profit from the development and usage of rapid tools able to certify meat and meat products with claims of quality, origin or species and rearing system, to guarantee that consumers are not being defrauded in the purchase (Prieto et al., 2017). The common analytical techniques currently available to the scope are destructive, time-consuming, and can only be employed at laboratory scale (Hassoun et al., 2020). Thus, researches are increasingly engaged in the development of rapid, and non-invasive methods to assess meat quality features (Kim et al., 2017; Prieto et al., 2017). In this regard, spectroscopic techniques are promising, being by nature non-destructive, environmentally friendly, fast, and easy to use (Squeo et al., 2019). Among them, previous studies have shown the feasibility of Near Infra-Red Spectroscopy (NIRS), coupled with chemometrics techniques, to: i) predict the chemical composition of meat (Barragán-Hernández et al., 2020; Dias et al., 2021; Fan et al., 2018; Ortiz et al., 2020); ii) predict the technological parameters and sensory attributes (Barbon et al., 2018; Savoia al., 2020; Wang et al., 2018); iii) predict carcass fat and meat quality (Fan et al., 2018); iv) classify and identify meat and meat products (López-Maestresalas et al., 2019; Weng et al., 2020). The spectral features of meat in the NIR range (800-2500 nm) depend on the overtones and combination bands of C–O, C–C, C–H, O–H, and N–H functional groups (Khaled et al., 2021; Prieto et al., 2017), as related to fat, water, and proteins (Barbin et al., 2013). Hence, the NIR spectrum provides information about the chemical composition of the material (Xiaobo et al., 2010), so that changes in the chemical constituents can form detailed fingerprints for quality detection (Prieto et al., 2009).

As for the meat authenticity, NIRS has found application for the evaluation of the dietary pattern (Arce et al., 2009; Dian et al., 2008; García-Olmo et al., 2009; Huang et al., 2015; Prieto et al., 2015; Zamora-Rojas et al., 2012) and the geographical origin of animals (Erasmus et al., 2016; Monahan et al., 2018; Sun et al., 2012). However, these studies faced the problem with a discriminant approach. Conversely, in our study a class-modelling approach has been used. Class-modelling, also recognised as one-class classification, is a family of multivariate supervised classification methods specifically developed to face problems in which only one class, i.e., the target class, is the focus and is recognised as the most suitable chemometric approach for food authentication (Forina et al., 2008; Rodionova et al., 2016a; Rodionova et al., 2016b; Oliveri, 2017). In this study, the target class was the fat material from a local pig breed reared extensively, known as *Apulo-Calabrese*, listed in the national herd book that is used as raw material in several Italian Protected Designation of Origin (PDO) meat products such as *Salsiccia di Calabria* (Registration Number PDO-IT-1568), *Soppressata di Calabria* (Registration Number PDO-IT-1569), *Capocollo di Calabria* (Registration Number PDO-IT-1570) and *Pancetta di Calabria* (Registration Number PDO-IT-1568) (https://ec.europa.eu). The fat portion has been chosen as possible authenticity marker because the scientific literature confirms that major differences between intensive and extensive rearing systems could be observed in the fatty acid composition of the fat portion (Estévez et al., 2006; Tejerina et al., 2012), and the latter is also influenced by the pig breed (De Smet et al., 2004; Huang et al., 2020).

The Data Driven variant of Soft Independent Modelling of Class Analogy (DD-SIMCA) (Rodionova et al., 2016a) was used as supervised class modelling approach.

2. Materials and methods

2.1 Sample set

A total set of 61 different section of subcutaneous fat were sampled from the local farm Salumi Martina Franca S.r.l. (Martina Franca, Italy). Forty-six (46) samples were from *Apulo-Calabrese* black pig breed reared extensively in “Massseria Pezze Mammarelle” woodland, located in Martina Franca, Taranto, Apulia (South Italy). The pigs were fed with natural plants such as acorns (*Quercus trojana* Webb), pasture grasses and wild fruits with the integration of cereal and legume-based feed (barley, field bean and pea) purchased from the feed mill Quarato (Noci, Bari, Apulia Ataly), and slaughtered at 180-200 kg. Fifteen (15) samples were from crossbreed F1 *Landrace* X *Large White* pigs purchased from a local slaughterhouse, reared intensively in Apulia and Basilicata (South Italy), and slaughtered at 160-170 kg. The sampling has been performed in six different months, encompassing two years (2021 and 2022) (Table 1).

2.2 NIR spectra acquisition

NIR spectra of fat samples were collected in reflectance mode by the integrating sphere module of the FT-NIR spectrophotometer (Nicolet iS50, Thermo Fisher Scientific, Waltham, MA, USA). Fat samples were filled in a sample cup (5 × 1 cm) that allows multi-sampling by rotating the cup during spectral acquisition. The acquisition conditions were: 4000-12500 cm^{-1} spectral range, 16 cm^{-1} resolution, 32 number of scans. Four spectra per each sample were collected at room temperature, recording a new background (32 scans) after every sample (4 spectra). A dark correction was applied to exclude the contribution of reflected light from the sampling window. The equipment was controlled by OMNIC software (Thermo Fisher Scientific Inc., Waltham, MA, USA). The full spectrum was used for further data elaboration.

2.3 Data Analysis

The NIR dataset was exported in “.csv” format and imported in MATLAB (R2021a, The MathWorks, Inc. MA, USA) for subsequent elaborations. First, the spectral replicates per each sample were averaged giving a final dataset of 61 × 1102 (samples × variables). Principal Component Analysis (PCA) was used as an unsupervised pattern recognition technique for data exploration, according to the model $X = TP^t + E$, where X is the original data matrix, T and P are the scores and loading matrices, respectively, and E is the matrix of residuals (Bro and Smilde, 2014). Different pre-processing were tested to remove unwanted signal variation (Oliveri et al., 2019; Rinnan et al., 2009), namely Standard Normal Variate (SNV; according to $x_{corr} = (x_{orig} - x_m)/s_d$, where x_{corr} is the corrected spectrum, x_{orig} is the original one, x_m is the mean spectrum, and s_d is the standard deviation of the sample spectrum), Multiplicative Scatter Correction (MSC; according to $x_{corr} = (x_{orig} - b_0)/b_{ref,1}$, where b_0 and $b_{ref,1}$ are the correction parameters from the first order polynomial using mean spectrum as reference), first derivative (1D, 11 data-point window, 3-order polynomial), and second derivative (2D, 11 data-point window, 3-order polynomial). Mean centring was used in all the cases as column scaling before performing PCA (Bro and Smilde, 2014). After exploration, DD-SIMCA was used as supervised class modelling approach (Rodionova et al., 2016a). In DD-SIMCA, after a decomposition step by PCA ($X = TP^t + E$), a score distance and an orthogonal distance are calculated per each sample, from which the total distance metric could be estimated and used to generate the decision rule (Zontov et al., 2017). Before class modelling, the dataset was split in calibration and validation set. The splitting scheme followed the so-called "rigorous" approach to one-class classification (Rodionova et al., 2016a). According to the rigorous approach, SIMCA models have been developed only by using samples belonging from the target class. In our case, the target class, i.e., the class whose authenticity we want to safeguard, is represented by the fat samples from *Apulo-Calabrese* extensive reared black pigs. Thus, the calibration set was represented by a portion of these fat samples (E class, n = 33) while the remaining (n = 13) were used for validation purpose. An α value of 0.001 was set for the model development, which represents the type I error, i.e., the rate of wrong rejection of target samples from the target class (Rodionova et al., 2016a). In the authentication context, α should be as lower as possible to reduce the probability that genuine samples from the target class are rejected from the model (Forina et al., 2008). Different models were built, and compared, considering i) a different number of principal components and ii) the mentioned spectral pre-processing (SNV, MSC, 1D, and 2D). The performances of the models were evaluated by assessing model's sensitivity (Sn) both for the training set (SnT) and for the prediction set (SnP). Sn, is defined as the true positive rate of a model, i.e., represents the number of objects of the target class correctly included in that class (Ballabio et al., 2018). The higher the Sn, the better the model. In this study, there are also samples from the alien class, i.e., fat from intensive rearing, which allow us to test the model's specificity (Sp). Sp, indicates the true negative rate, i.e., the number of objects from the alien class correctly rejected from the target class model (Ballabio et al., 2018). Sn and Sp have been calculated as reported in Zontov et al. (2017). Data elaboration was carried out in MATLAB, by using original codes, the PLS_toolbox (Eigenvector Research Inc., USA) and the DD-SIMCA toolbox (Teixeira et al., 2021; Zontov et al., 2017).

3. Results and Discussion

3.1 NIR spectra and data exploration

Figure 1 depicts the raw mean NIR spectra of the fat classes under investigation. The spectra are similar to those reported for other pig meat-based samples (Mabood et al., 2020; Varrà et al., 2020; Zamora-Rojas et al., 2012) and could be roughly divided into three main regions. The first one, between 4000 and 6000 cm^{-1}, shows different absorption bands (at 4258, 4327, 4666, 5176, 5677, and 5785 cm^{-1}) associated to the combination and the first overtone bands of the CH, CH_2, CH_3, CC, CO, and NH_2 functional groups, and H_2O. The second region is characterized by two broad bands at around 6950 cm^{-1}, with a shoulder at 7181 cm^{-1}, and at 8254 cm^{-1}, with a shoulder at 8554 cm^{-1}, that could be associated mainly to the CH, CH_2, CH_3 first and second overtones, respectively. Finally, two little bands were observed in the third region, at around 9596 and 10745 cm^{-1}, that could be linked to the second overtone of NH and the third overtone of CH, CH_2, CH_3 functional groups, respectively (Dixit et al., 2017; Xiaobo et al., 2010).

Being the samples under study constituted by fat, it was expected that the absorptions in the NIR range were linked mainly to the CH groups, that represents the core of the lipid structure. Considering the classes, no important differences were observable in the raw spectra, confirming the necessity of chemometric tools to extract useful information from those data (Squeo et al., 2019). So, PCA was used for data exploration. Figure 2 reports the results of the PCA carried out on the SNV pre-processed data matrix. The results obtained by applying other pre-processing can be observed in the supplementary material. The first two principal components (PCs) explained more than 90% of the total variability. From the score plot (Figure 2A), it could be observed that PC1 did not highlight information related to the rearing system and the samples from the two classes overlapped. Differently, along PC2 it could be observed a preferable distribution of the I class samples that were roughly positioned on the negative half of PC2, although there was no clear separation with E samples. The inspection of other PCs (data not shown) did not highlight useful information to our scope. The loading plot (Figure 2B) showed the variables responsible for this displacement of the I class samples, which were those having higher and negative loadings on PC2, mostly those in the regions 4000-4300 cm^{-1}, 5000-5300 cm^{-1}, 6500-7000 cm^{-1}. As illustrated above, these spectral regions correspond to those of the combination and the first and second overtone bands of CH, OH, and CO (Xiaobo et al., 2010), which are linked to the fatty acids. The same conclusions drawn for the SNV pre-processed data could be outlined considering the MSC pre-processing as can be observed in Figure S1. Differently, the application of the 1D and 2D (Figures S2 and S3, respectively) brought to a progressively increasing separation of the samples according to the classes in the space of the first two PCs, which is almost net in the case of 2D. In light of these results, the derivative pre-processing seemed to be the most suitable to achieve discrimination. Nonetheless, it should be recalled that our problem is not to discriminate these two classes but to define a class model useful for the authentication of the target class, E.

3.2 Class-modelling results

3.2.1 Models calibration

As previously stated, in the context of authentication problems, the most suitable chemometric approach is to define a class model for the target class that will be used to classify unknown samples as members of such class or not. SIMCA, in its different variants, is one of the simpler and more powerful class modelling tools. In this study, the DD variant was used (Zontov et al., 2017) and Table 2 reports the obtained results. Figure 3 presents graphically the evolution of SnT, SnP, and Sp for an increasing number of PCs and for the different pre-processing. All the models were calculated by setting an α value of 0.001, which corresponds to defining a 99.90% confidence level for the target class. As reported in Forina et al. (2008), reducing the α value, and thus increasing the confidence level for the class and the model sensitivity, is often mandatory in the food authentication context because of the need to not exclude genuine samples of the target class from the model. By setting the value at 0.001, the accepted type I error is reduced to 0.1%. On the other hand, it should be considered that, generally, reducing the α value causes an increase in the β value (Rodionova et al., 2016a), which means that a higher proportion of alien samples could be classified as target samples, i.e., a higher amount of false positive. The results of the calibration showed that an excellent sensitivity (SnT) was obtained using all the pre-processing tools, regardless of the number of PCs. Only in the case of SNV and MSC, one extreme sample was identified when 6 PCs were used (Table 2, Figure 3). These excellent results in calibration were expected considering the chosen α value, as previously commented. However, on the basis of these results only, it would be a hard task to decide which model to choose, both in terms of complexity and pre-processing. Hence, it is mandatory to test the models to obtain an insight about their predictive performances, as discussed in the next section.

3.2.2 Models validation

The calibrated models were tested against a validation set of 13 random chosen E class samples and, finally also against a set of 15 I class samples to calculate the β value and the Sp.

Considering SNV pre-processing, the results showed an excellent SnP from 1 up to 6 PCs, after which a progressive decreasing in the SnP was observed, corresponding to an increasing number of samples not recognized as members of the target class (Table 2). On the other hand, Sp was extremely low with 1 PC but suddenly increased by considering more PCs. According to SnT and SnP, the models from 1 to 5 PCs were equivalently good. However, by considering also the Sp, the model with 4 PCs was the best compromise. MSC performed worse than SNV, in contrast with another similar study (Fernández-Barroso et al., 2021). In fact, SnP suddenly decreased from 2 PCs onward (Table 2, Figure 3). Conversely, MSC allowed to reach an outstanding Sp starting from the model with 2 PCs (Figure 3) with probabilities of type II error equal to 0 (Table 2). 1D and 2D models showed similar performances. In both cases, up to a reduced number of PCs (3 and 4, respectively) the models showed excellent SnP. Afterwards, the SnP started decreasing, although not in a dramatic manner, while Sp increased. For these models, the optimal dimensionality, chosen on the basis of SnT and SnP, brought to models with generally low Sp and β values equal to 0.37 and 0.24, respectively. Overall, in all the cases, an inverse relationship between SnP and PCs has been observed. As commented by Rodionova et al. (2016b), this result clearly reflects the possibility of overfitting of the model to the training dataset.

3.2.3 Optimal class model selection

A major issue when developing a class model regards model optimization, i.e., the definition of the optimal number of PCs, which is of crucial importance (Rodionova et al., 2016b). According to the “rigorous” class modelling approach, the best model should be chosen only on the basis of SnT and SnP. Following this approach, in this study different models are found to be equivalently good. However, when the classes under study tend to overlap, such as in this case (Figure 2A), it might help considering also alien objects if available (Rodionova et al., 2016a). In this perspective, our results suggest that the best trade-off is obtained with SNV pre-processing and 4 PCs. In this case, SnT, SnP, and Sn were 100% with a calculated β value equal to 10% (Table 2). Figure 4 shows the acceptance plots obtained from this model (SNV, 4 PCs) for the training set, the target class validation set, and the alien class validation set, respectively. In line with our outcomes, other recent studies in different fields presented interesting results obtained combining NIR spectroscopy and DD-SIMCA to solve authentication problems. Teixeira et al. (2022) in the classification of authentic and adulterated samples of goat dairy beverage, obtained an excellent predictive performance with 95% (2 PCs; $\alpha = 0.01$) and 100% (1 PCs; $\alpha = 0.01$) sensitivity value, using the full spectral range and selected variables, respectively. Rodionova et al. (2016b) used this approach to successfully trace the authenticity of olive samples, while Mazivila et al. (2020) used DD-SIMCA for the authenticity of cow milk powders.

4. Conclusions

The results obtained in this work proved that NIR spectroscopy coupled with DD-SIMCA class modelling is a powerful analytical approach when testing the authenticity of pork fat according to the rearing system. This analytical methodology has the advantage of being naturally clean, rapid, and non-destructive, matching the contemporary requirements for fast and environmentally-friendly analytical methods. The NIR signal gave chemical information about the samples under study and, although the spectral signals of the considered classes were very similar, by a proper DD-SIMCA model tuning (i.e., definition of the optimal number of PCs and spectral pre-processing), outstanding performances in classification were obtained. These promising results will be confirmed with an increased dataset and by an intense validation. In this study, the feasibility of NIR-DD-SIMCA was tested on the sole fat portion. Thus, an exciting perspective will be the application of the proposed method for the authentication of more complex meat products from extensive rearing.

Acknowledgements

The authors are grateful to Francesco Carriero of “Salumi Martina Franca S.r.l.” (Martina Franca, Italy) for furnishing the fat samples.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Arce, L., Domínguez-Vidal, A., Rodríguez-Estévez, V., López-Vidal, S., Ayora-Cañada, M. J., Valcárcel, M. (2009). Feasibility study on the use of infrared spectroscopy for the direct authentication of Iberian pig fattening diet. *Analytica Chimica Acta*, *636*, 183-189.

Ballabio, D., Grisoni, F., Todeschini, R. (2018). Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems*, *174*, 33-44.

Barbin, D. F., ElMasry, G., Sun, D. W., Allen, P. (2013). Non-destructive determination of chemical composition in intact and minced pork using near-infrared hyperspectral imaging. *Food Chemistry, 138*, 1162-1171.

Barbon, S., Costa Barbon, A. P. A. D., Mantovani, R. G., Barbin, D. F. (2018). Machine learning applied to near-infrared spectra for chicken meat classification. *Journal of Spectroscopy, 2018*, 8949741.

Barragán-Hernández, W., Mahecha-Ledesma, L., Burgos-Paz, W., Olivera-Angel, M., Angulo-Arizala, J. (2020). Using near-infrared spectroscopy to determine intramuscular fat and fatty acids of beef applying different prediction approaches. *Journal of Animal Science, 98*, skaa342.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, *6*, 2812-2831.

COUNCIL REGULATION (EC) No 1099/2009. (2009) *Official Journal of the European Union*, *L303*, 1-30.

De Smet, S., Raes, K., Demeyer, D. (2004). Meat fatty acid composition as affected by fatness and genetic factors: a review. *Animal Research*, *53*, 81-98.

Dian, P. H. M., Andueza, D., Jestin, M., Prado, I. N., Prache, S. (2008). Comparison of visible and near infrared reflectance spectroscopy to discriminate between pasture-fed and concentrate-fed lamb carcasses. *Meat Science*, *80*, 1157-1164.

Dias, C. M., Nunes, H. P., Melo, T. M. M. V., Rosa, H. J. D., Silva, C. C. G., Borba, A. E. S. (2021). Application of Near Infrared Reflectance (NIR) spectroscopy to predict the moisture, protein, and fat content of beef for gourmet hamburger preparation. *Livestock Science*, *254*, 104772.

Dixit, Y., Casado-Gavalda, M. P., Cama-Moncunill, R., Cama-Moncunill, X., Markiewicz-Keszycka, M., Cullen, P. J., Sullivan, C. (2017). Developments and challenges in online NIR spectroscopy for meat processing. *Comprehensive Reviews in Food Science and Food Safety*, *16*, 1172-1187.

Erasmus, S. W., Manley, M., Muller, M., Hoffman, L. C. (2016). MicroNIR spectroscopy for the authentication of south African lamb. In *Proceedings of 62nd international congress of meat science and technology: Meat for global sustainability* (p.36). Bangkok, Thailand.

Estévez, M., Morcuende, D., Cava, R. (2006). Extensively reared Iberian pigs versus intensively reared white pigs for the manufacture of frankfurters. *Meat Science*, *72*, 356-364.

European Union (2022). Annual Report Alert and Cooperation Netwprk. Luxembourg: Publication Office 250 of the European Union, 2022. Accessed: July 7th, 2022.

Fan, Y., Liao, Y., Cheng, F. (2018). Predicting of intramuscular fat content in pork using near infrared spectroscopy and multivariate analysis. *International Journal of Food Properties*, *21*, 1180-1189.

Farm Animal Welfare Council. (1992). FAWC updates the five freedoms. *Veterinary Record*, *17*, 357.

Fernández-Barroso, M. Á., Parrini, S., Muñoz, M., Palma-Granados, P., Matos, G., Ramírez, L., Bozzi, R. (2021). Use of NIRS for the assessment of meat quality traits in open-air free-range Iberian pigs. *Journal of Food Composition and Analysis*, *102*, 104018.

Forina, M., Oliveri, P., Lanteri, S., Casale, M. (2008). Class-modeling techniques, classic and new, for old and new problems. *Chemometrics and Intelligent Laboratory Syste*ms, 93, 132-148.

García-Olmo, J., Garrido-Varo, A., De Pedro, E. (2009). Classification of real farm conditions Iberian pigs according to the feeding regime with multivariate models developed by using fatty acids composition or NIR spectral data. *Grasas y Aceites*, *60*, 233-237.

Hassoun, A., Måge, I., Schmidt, W. F., Temiz, H. T., Li, L., Kim, H. Y., Cozzolino, D. (2020). Fraud in animal origin food products: Advances in emerging spectroscopic detection methods over the past five years. *Foods*, *9*, 1069.

Huang, Y., Andueza, D., de Oliveira, L., Zawadzki, F., Prache, S. (2015). Comparison of visible and near infrared reflectance spectroscopy on fat to authenticate dietary history of lambs. *Animal*, *9*, 1912-1920.

Huang, Y., Zhou, L., Zhang, J., Liu, X., Zhang, Y., Cai, L., Huang, L. (2020). A large-scale comparison of meat quality and intramuscular fatty acid composition among three Chinese indigenous pig breeds. *Meat Science*, *168*, 108182.

Khaled, A. Y., Parrish, C. A., Adedeji, A. (2021). Emerging nondestructive approaches for meat quality and safety evaluation—A review. *Comprehensive Reviews in Food Science and Food Safety*, *20*, 3438-3463.

Kim, J. S., Hwang, I. M., Lee, G. H., Park, Y. M., Choi, J. Y., Jamila, N., Kim, K. S. (2017). Geographical origin authentication of pork using multi-element and multivariate data analyses. *Meat Science*, *123*, 13-20.

López-Maestresalas, A., Insausti, K., Jarén, C., Pérez-Roncal, C., Urrutia, O., Beriain, M. J., Arazuri, S. (2019). Detection of minced lamb and beef fraud using NIR spectroscopy. *Food Control*, *98*, 465-473.

Mabood, F., Boqué, R., Alkindi, A. Y., Al-Harrasi, A., Al Amri, I. S., Boukra, S., Kadim, I. (2020). Fast detection and quantification of pork meat in other meats by reflectance FT-NIR spectroscopy and multivariate analysis. *Meat Science*, *163*, 108084.

Mazivila, S. J., Páscoa, R. N., Castro, R. C., Ribeiro, D. S., Santos, J. L. (2020). Detection of melamine and sucrose as adulterants in milk powder using near-infrared spectroscopy with DD-SIMCA as one-class classifier and MCR-ALS as a means to provide pure profiles of milk and of both adulterants with forensic evidence: A short communication. *Talanta*, *216*, 120937.

Monahan, F. J., Schmidt, O., Moloney, A. P. (2018). Meat provenance: Authentication of geographical origin and dietary background of meat. *Meat Science*, *144*, 2-14.

Oliveri, P. (2017). Class-modelling in food analytical chemistry: development, sampling, optimisation and validation issues–a tutorial. *Analytica Chimica Acta*, 982, 9-19.

Oliveri, P., Malegori, C., Simonetti, R., Casale, M. (2019). The impact of signal pre-processing on the final interpretation of analytical outcomes–A tutorial. *Analytica Chimica Acta*, *1058*, 9-17.

Ortiz, A., Parrini, S., Tejerina, D., Pinto de Araújo, J. P., Čandek-Potokar, M., Crovetti, A., Bozzi, R. (2020). Potential use of near-infrared spectroscopy to predict fatty acid profile of meat from different european autochthonous pig breeds. *Applied Sciences*, *10*, 5801.

Prieto, N., Juarez, M., Larsen, I. L., Lopez-Campos, O., Zijlstra, R. T., Aalhus, J. L. (2015). Rapid discrimination of enhanced quality pork by visible and near infrared spectroscopy. *Meat Science*, *110*, 76-84.

Prieto, N., Pawluczyk, O., Dugan, M. E. R., Aalhus, J. L. (2017). A review of the principles and applications of near-infrared spectroscopy to characterize meat, fat, and meat products. *Applied Spectroscopy, 71*, 1403-1426.

Prieto, N., Roehe, R., Lavín, P., Batten, G., Andrés, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Science*, *83*, 175-186.

Rinnan, Å., Van Den Berg, F., Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*, 1201-1222.

Rodionova, O. Y., Oliveri, P., Pomerantsev, A. L. (2016a). Rigorous and compliant approaches to one-class classification. *Chemometrics and Intelligent Laboratory Systems, 159,* 89-96.

Rodionova, O. Y., Titova, A. V., Pomerantsev, A. L. (2016b). Discriminant analysis is an inappropriate method of authentication. *TrAC Trends in Analytical Chemistry, 78,* 17-22.

Sajali, N., Wong, S. C., Abu Bakar, S., Khairil Mokhtar, N. F., Manaf, Y. N., Yuswan, M. H., Mohd Desa, M. N. (2021). Analytical approaches of meat authentication in food. *International Journal of Food Science & Technology*, *56*, 1535-1543.

Savoia, S., Albera, A., Brugiapaglia, A., Di Stasio, L., Ferragina, A., Cecchinato, A., Bittante, G. (2020). Prediction of meat quality traits in the abattoir using portable and hand-held near-infrared spectrometers. *Meat Science*, *161*, 108017.

Squeo, G., Grassi, S., Paradiso, V. M., Alamprese, C., Caponio, F. (2019). FT-IR extra virgin olive oil classification based on ethyl ester content. *Food Control*, *102*, 149-156.

Sun, S., Guo, B., Wei, Y., Fan, M. (2012). Classification of geographical origins and prediction of δ13C and δ15N values of lamb meat by near infrared reflectance spectroscopy. *Food Chemistry*, *135*, 508-514.

Teixeira, J. L. D. P., Caramês, E. T. D. S., Baptista, D. P., Gigante, M. L., Pallone, J. A. L. (2022). Adulteration detection in goat dairy beverage through NIR spectroscopy and DD-SIMCA. *Food Analytical Methods*, *15*, 783-791.

Tejerina, D., García-Torres, S., De Vaca, M. C., Vázquez, F. M., Cava, R. (2012). Effect of production system on physical–chemical, antioxidant and fatty acids composition of Longissimus dorsi and Serratus ventralis muscles from Iberian pig. *Food Chemistry*, *133*, 293-299.

Varrà, M. O., Fasolato, L., Serva, L., Ghidini, S., Novelli, E., Zanardi, E. (2020). Use of near infrared spectroscopy coupled with chemometrics for fast detection of irradiated dry fermented sausages. *Food Control*, *110*, 107009.

Wang, W., Peng, Y., Sun, H., Zheng, X., Wei, W. (2018). Real-time inspection of pork quality attributes using dual-band spectroscopy. *Journal of Food Engineering*, *237,* 103-109.

Weng, S., Guo, B., Tang, P., Yin, X., Pan, F., Zhao, J., Zhang, D. (2020). Rapid detection of adulteration of minced beef using Vis/NIR reflectance spectroscopy with multivariate methods. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *230*, 118005.

Xiaobo, Z., Jiewen, Z., Povey, M. J., Holmes, M., Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, *667*, 14-32.

Zamora-Rojas, E., Pérez-Marín, D., De Pedro-Sanz, E., Guerrero-Ginel, J. E., Garrido-Varo, A. (2012). In-situ Iberian pig carcass classification using a micro-electro-mechanical system (MEMS)-based near infrared (NIR) spectrometer. *Meat Science*, *90*, 636-642.

Zontov, Y. V., Rodionova, O. Y., Pomerantsev, A. L., Kucheryavskiy, S. V. (2017). DD-SIMCA–a MATLAB GUI tool for data driven SIMCA approach. *Chemometrics and Intelligent Laboratory Systems*, *167,* 23-28.

Figures captions

Figure 1. Mean NIR raw spectra of fat samples from extensive reared *Apulo-Calabrese* black pig (E; black line) and intensive reared pink pig (I; red line).

Figure 2. Score plot (A) and loading plot (B) of the Principal Component Analysis carried out on the SNV pre-processed data matrix. E, extensive reared *Apulo-Calabrese* black pig; I, intensive reared pink pig.

Figure 3. Graphical representation the evolution of SnT (sensitivity calculated on the training set), SnP (sensitivity calculated on the target class test set), and Sp (specificity calculated on the alien class) for an increasing number of PCs and for the different pre-processing (SNV, standard normal variate; MSC, multiplicative scatter correction; 1D, first derivative; 2D, second derivative).

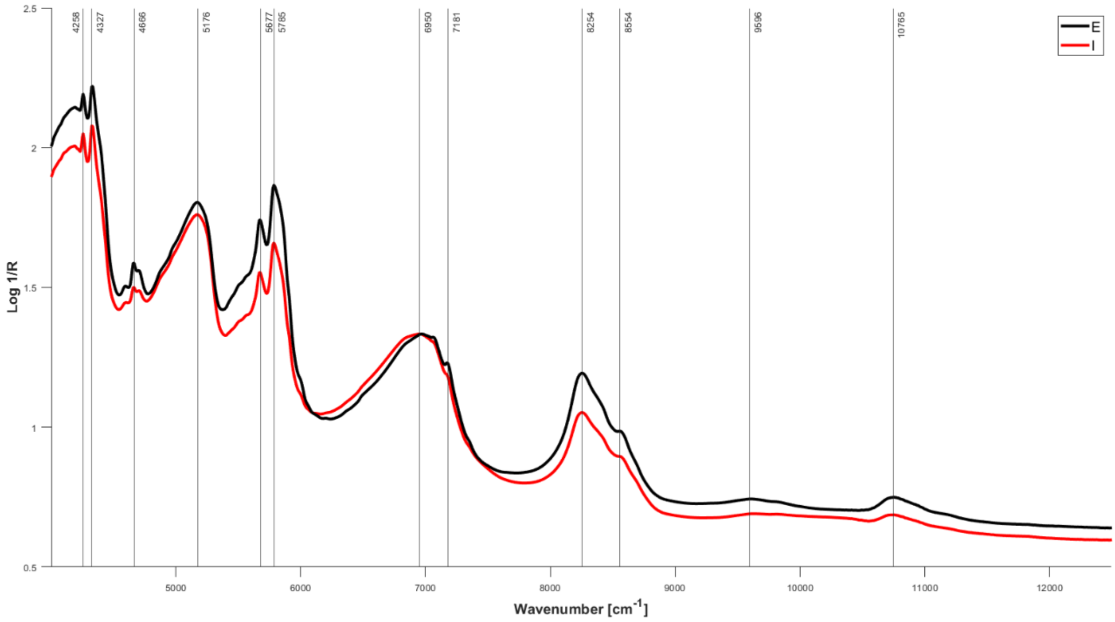
Figure 4. Acceptance plots obtained from the model (SNV, 4 PCs) for A) the training set, B) the target class validation set, and C) the alien class validation set.

Supplementary material

Figure S1. Score plot (A) and loading plot (B) of the Principal Component Analysis carried out on the MSC pre-processed data matrix. E, extensive reared *Apulo-Calabrese* black pig; I, intensive reared pink pig.

Figure S2. Score plot (A) and loading plot (B) of the Principal Component Analysis carried out on the 1D pre-processed data matrix. E, extensive reared *Apulo-Calabrese* black pig; I, intensive reared pink pig.

Figure S3. Score plot (A) and loading plot (B) of the Principal Component Analysis carried out on the 2D pre-processed data matrix. E, extensive reared *Apulo-Calabrese* black pig; I, intensive reared pink pig.



2.5
2
1.5
1
0.5
Log 1/R
4258
4327
4666
5176
5677
5785
6950
7181
8254
8554
9596
10765
E
I
5000
6000
7000
8000
9000
10000
11000
12000
Wavenumber [cm-1]

Figure 1.

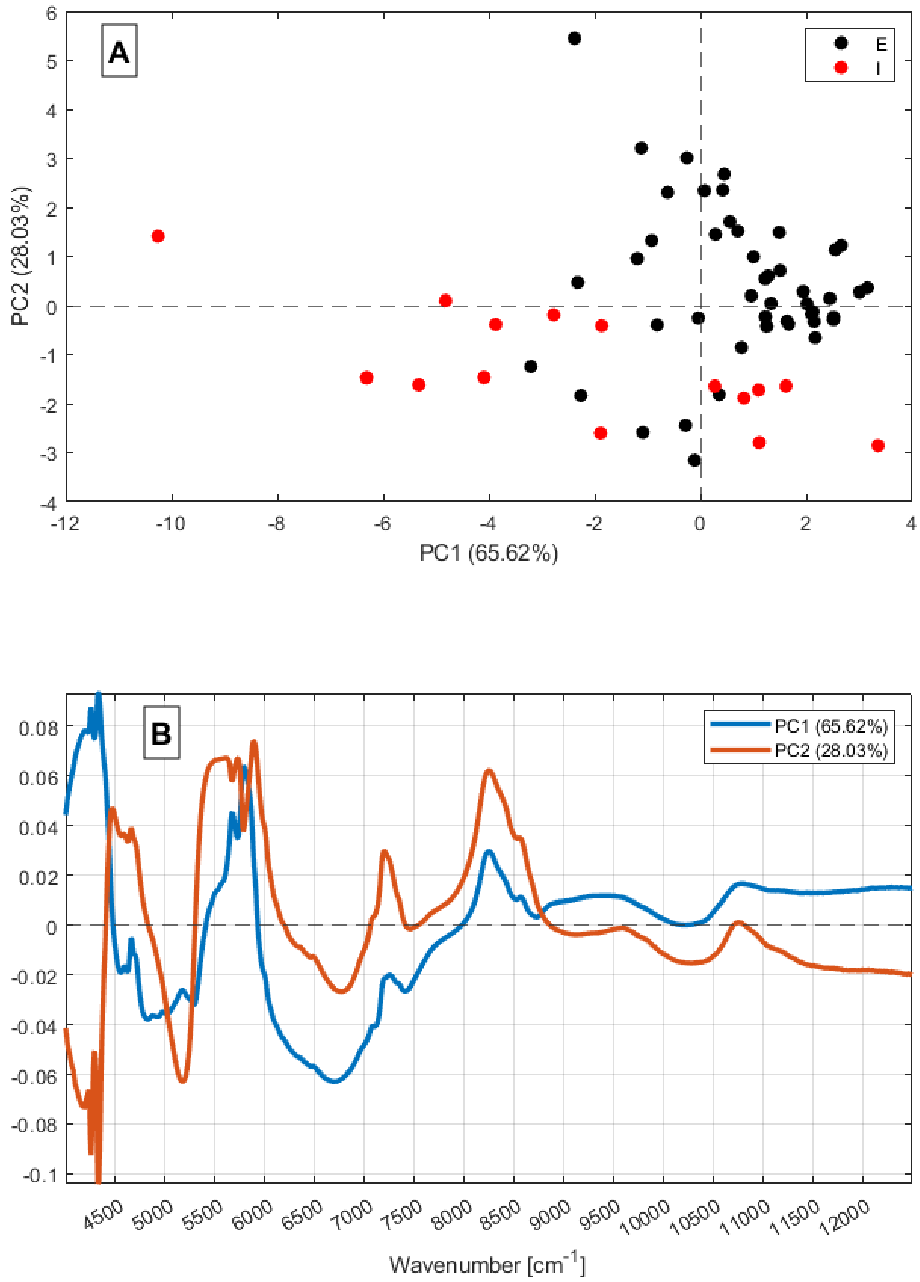


Figure 2.

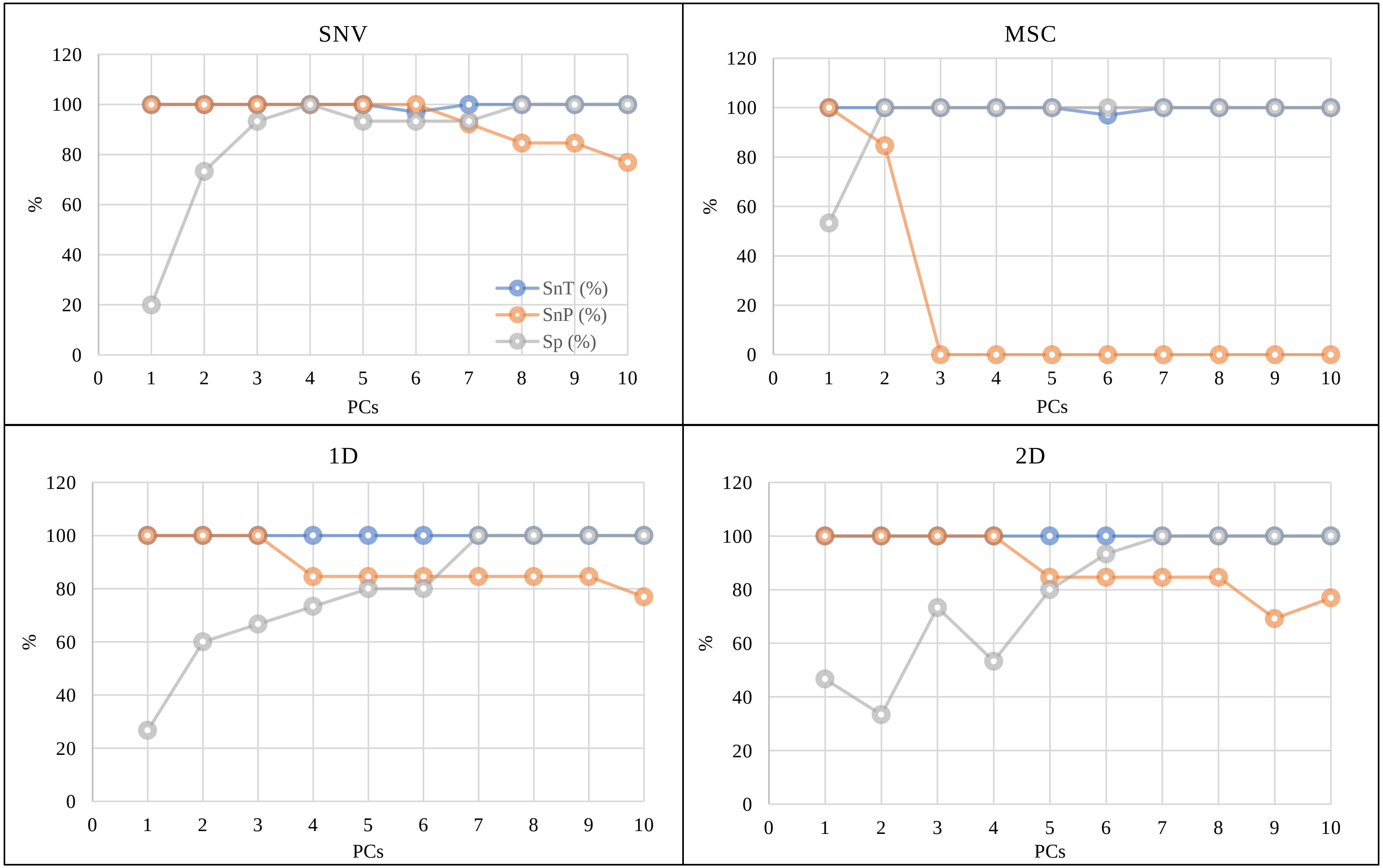


Figure 3.

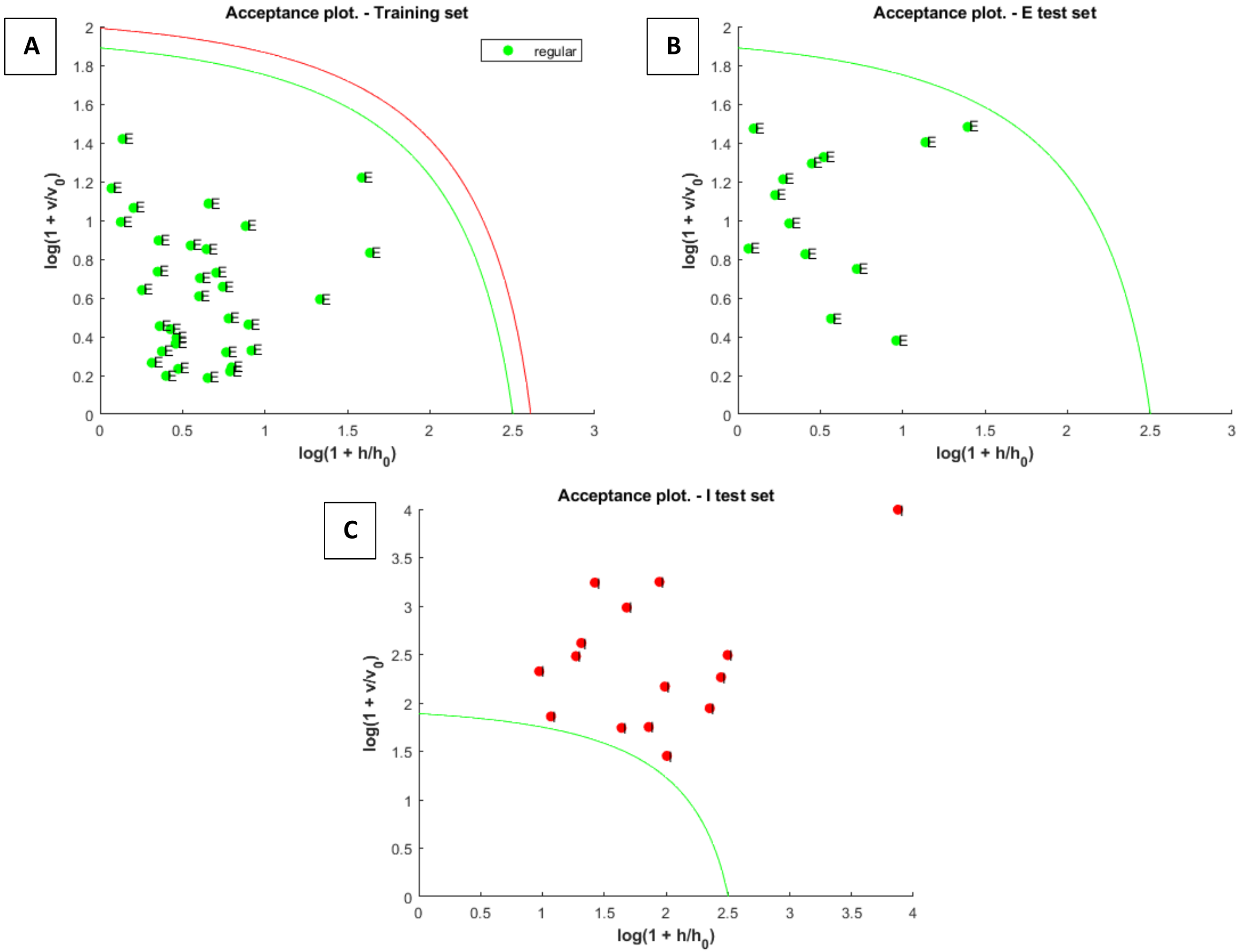

A
Acceptance plot. - Training set
regular
log(1 + v/v0)
log(1 + h/h0)
B
Acceptance plot. - E test set
log(1 + v/v0)
log(1 + h/h0)
C
Acceptance plot. - I test set
log(1 + v/v0)
log(1 + h/h0)

Figure 4.

Table 1. Presentation of the samples.

No. of independent samples	Period of sampling	Pig breed	Rearing system
4	May ‘21	*Landrace* X *Large White*	Intensive
5	May ‘21	*Apulo-Calabrese*	Extensive free-range
6	June ‘21	*Landrace* X *Large White*	Intensive
10	June ‘21	*Apulo-Calabrese*	Extensive free-range
5	June ‘21	*Apulo-Calabrese*	Extensive free-range
4	July ‘21	*Apulo-Calabrese*	Extensive free-range
2	January ‘22	*Landrace* X *Large White*	Intensive
10	January ‘22	*Apulo-Calabrese*	Extensive free-range
3	March ‘22	*Landrace* X *Large White*	Intensive
4	March ‘22	*Apulo-Calabrese*	Extensive free-range
4	April ‘22	*Apulo-Calabrese*	Extensive free-range
4	April ‘22	*Apulo-Calabrese*	Extensive free-range

Table 2. Results of DD-SIMCA.

			Calibration TARGET CLASS				Prediction TARGET CLASS			Prediction ALIEN CLASS			
Preprocessing	PCs	α	Samples	Extremes	Outliers	SnT (%)	Samples	External	SnP (%)	Samples	External	β	Sp (%)
SVN + MC	1	0.001	33	0	0	100.00	13	0	100.00	15	3	0.89	20.00
	2	0.001	33	0	0	100.00	13	0	100.00	15	11	0.29	73.33
	3	0.001	33	0	0	100.00	13	0	100.00	15	14	0.04	93.33
	4	0.001	33	0	0	100.00	13	0	100.00	15	15	0.10	100.00
	5	0.001	33	0	0	100.00	13	0	100.00	15	14	0.15	93.33
	6	0.001	33	1	0	96.97	13	0	100.00	15	14	0.05	93.33
	7	0.001	33	0	0	100.00	13	1	92.31	15	14	0.02	93.33
	8	0.001	33	0	0	100.00	13	2	84.62	15	15	0.01	100.00
	9	0.001	33	0	0	100.00	13	2	84.62	15	15	0.01	100.00
	10	0.001	33	0	0	100.00	13	3	76.92	15	15	0.00	100.00
MSC + MC	1	0.001	33	0	0	100.00	13	0	100.00	15	8	0.44	53.33
	2	0.001	33	0	0	100.00	13	2	84.62	15	15	0.00	100.00
	3	0.001	33	0	0	100.00	13	13	0.00	15	15	0.00	100.00
	4	0.001	33	0	0	100.00	13	13	0.00	15	15	0.00	100.00
	5	0.001	33	0	0	100.00	13	13	0.00	15	15	0.00	100.00
	6	0.001	33	1	0	96.97	13	13	0.00	15	15	0.00	100.00
	7	0.001	33	0	0	100.00	13	13	0.00	15	15	0.00	100.00
	8	0.001	33	0	0	100.00	13	13	0.00	15	15	0.00	100.00
	9	0.001	33	0	0	100.00	13	13	0.00	15	15	0.00	100.00
	10	0.001	33	0	0	100.00	13	13	0.00	15	15	0.00	100.00
1D + MC	1	0.001	33	0	0	100.00	13	0	100.00	15	4	0.65	26.67
	2	0.001	33	0	0	100.00	13	0	100.00	15	9	0.44	60.00
	3	0.001	33	0	0	100.00	13	0	100.00	15	10	0.37	66.67
	4	0.001	33	0	0	100.00	13	2	84.62	15	11	0.19	73.33
	5	0.001	33	0	0	100.00	13	2	84.62	15	12	0.17	80.00
	6	0.001	33	0	0	100.00	13	2	84.62	15	12	0.19	80.00
	7	0.001	33	0	0	100.00	13	2	84.62	15	15	0.02	100.00
	8	0.001	33	0	0	100.00	13	2	84.62	15	15	0.03	100.00

	9	0.001	33	0	0	100.00	13	2	84.62	15	15	0.02	100.00
	10	0.001	33	0	0	100.00	13	3	76.92	15	15	0.01	100.00
	1	0.001	33	0	0	100.00	13	0	100.00	15	7	0.54	46.67
	2	0.001	33	0	0	100.00	13	0	100.00	15	5	0.60	33.33
2D + MC	3	0.001	33	0	0	100.00	13	0	100.00	15	11	0.24	73.33
	4	0.001	33	0	0	100.00	13	0	100.00	15	8	0.37	53.33
	5	0.001	33	0	0	100.00	13	2	84.62	15	12	0.23	80.00
	6	0.001	33	0	0	100.00	13	2	84.62	15	14	0.16	93.33
	7	0.001	33	0	0	100.00	13	2	84.62	15	15	0.05	100.00
	8	0.001	33	0	0	100.00	13	2	84.62	15	15	0.01	100.00
	9	0.001	33	0	0	100.00	13	4	69.23	15	15	0.01	100.00
	10	0.001	33	0	0	100.00	13	3	76.92	15	15	0.01	100.00

PCs, number of principal components; α, value set for the type I error; SnT, sensitivity calculated on the training set; SnP, sensitivity calculated on the target class test set; β, calculated value of the type II error; Sp, specificity calculated on the alien class; SNV, standard normal variate; MSC, multiplicative scatter correction; 1D, first derivative; 2D, second derivative; MC, mean centering.

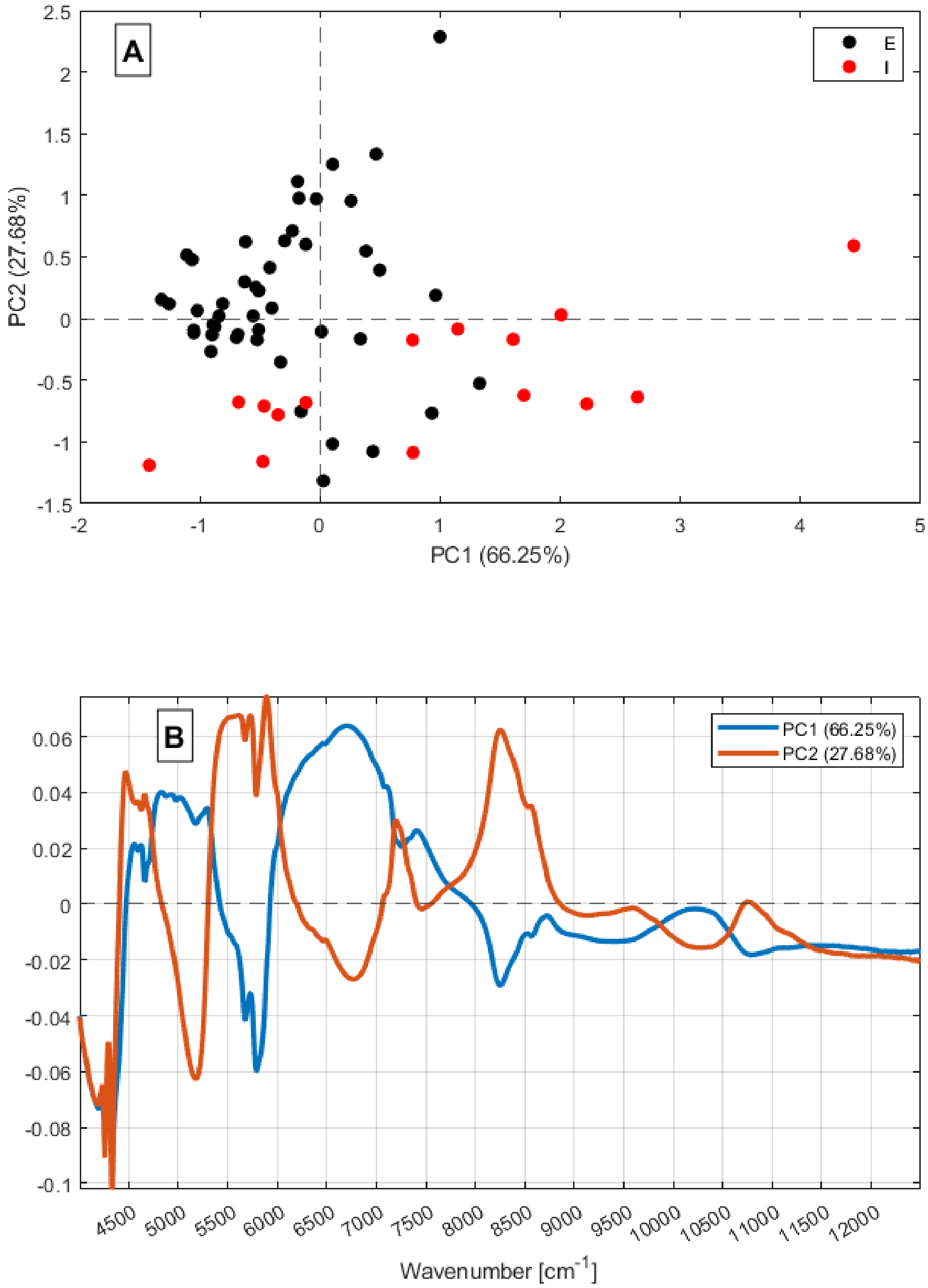


Figure S1.

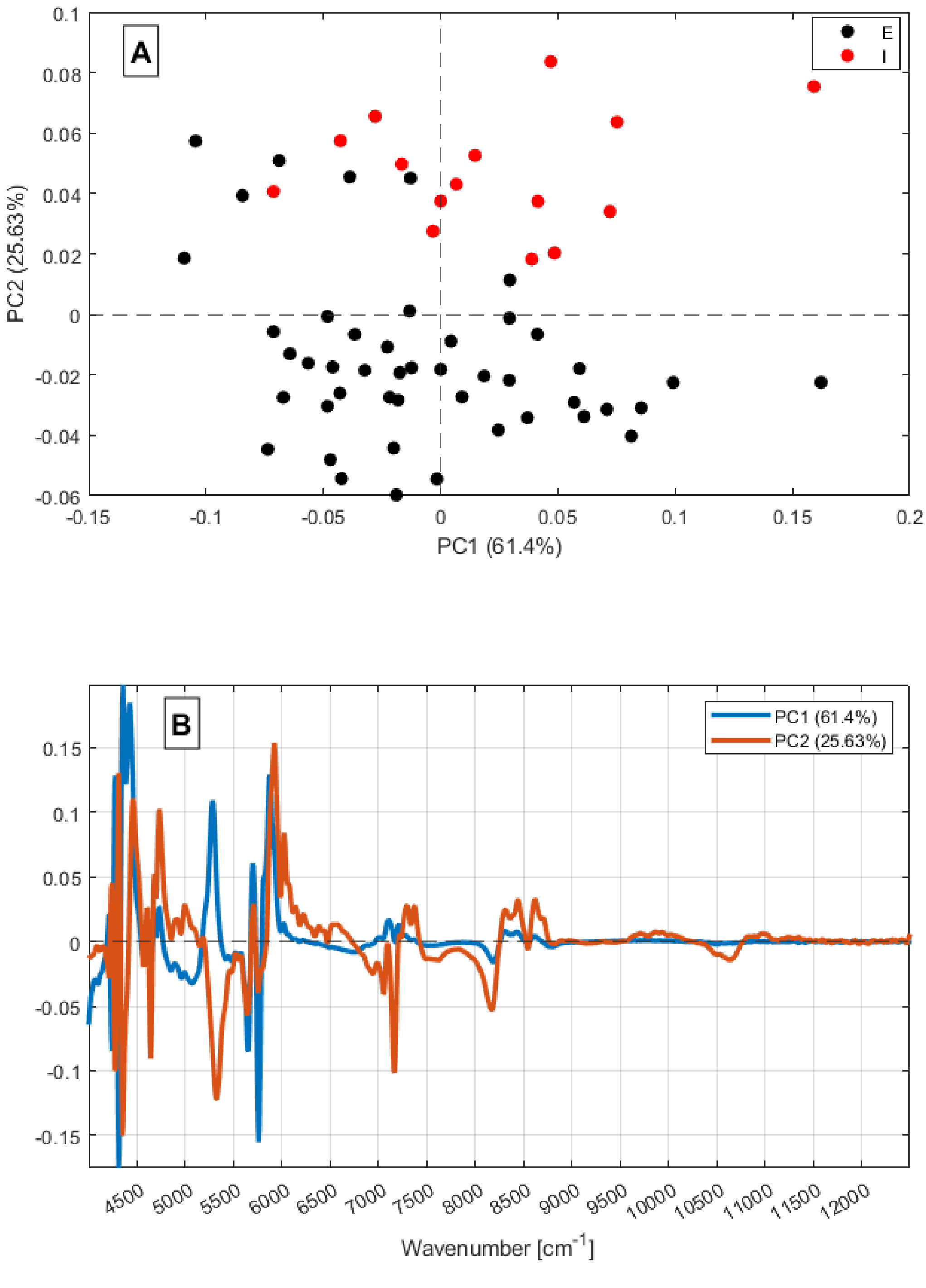
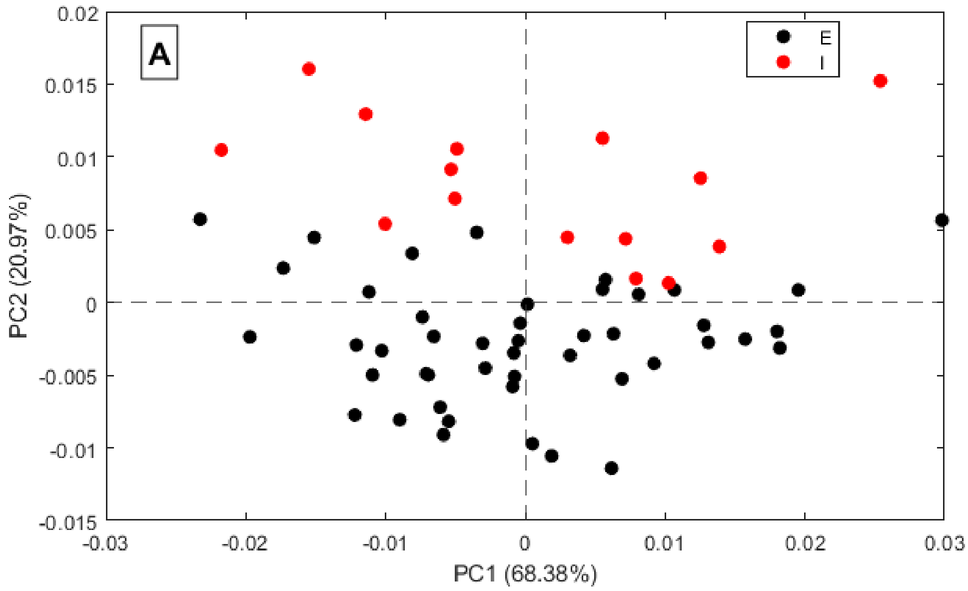


Figure S2.


A
E
I
PC2 (20.97%)
PC1 (68.38%)
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-0.03
-0.02
-0.01
0
0.01
0.02
0.03

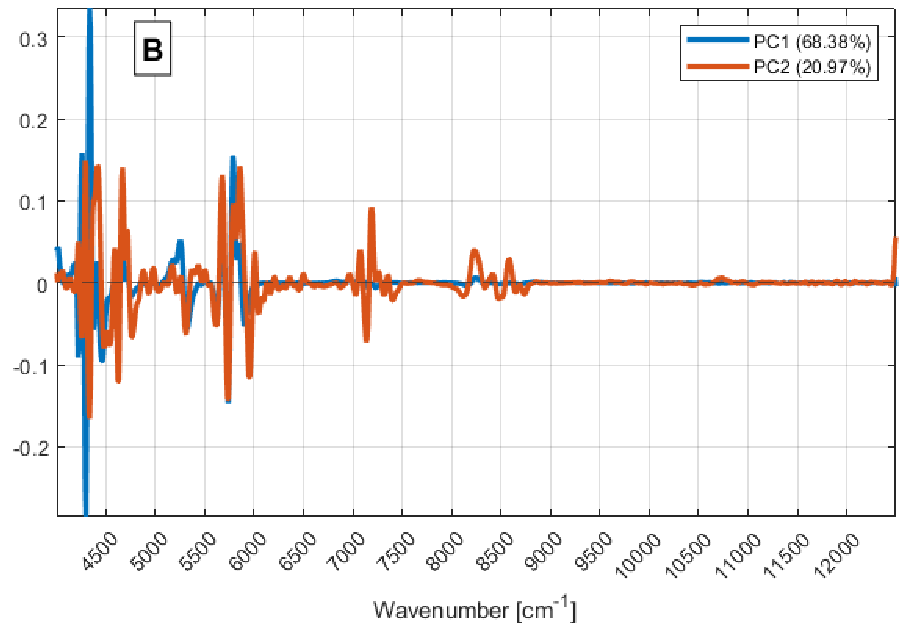

B
PC1 (68.38%)
PC2 (20.97%)
0.3
0.2
0.1
0
-0.1
-0.2
4500
5000
5500
6000
6500
7000
7500
8000
8500
9000
9500
10000
10500
11000
11500
12000
Wavenumber [cm-1]

Figure S3.

Highlights

- NIR spectra of fat from different pigs were recorded for authenticity purpose
- PCA showed that the classes overlapped
- DD-SIMCA class modelling allows to define a proper target class model
- By proper model tuning, excellent sensitivity and specificity were obtained
- New perspective for the authenticity check of pork meat products

CRediT authorship contribution statement

Michela Pia Totaro: Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing; **Giacomo Squeo**: Methodology, Software, Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing; **Davide De Angelis**: Investigation, Data Curation, Writing - Review & Editing; **Antonella Pasqualone**: Resources, Writing - Review & Editing; **Francesco Caponio**: Resources, Writing - Review & Editing; **Carmine Summo**: Conceptualization, Methodology, Writing - Review & Editing, Supervision.