Probabilistic classification models for the *in situ* authentication of Iberian pig carcasses using Near Infrared Spectroscopy

Dolores Pérez-Marín[a*], Tom Fearn[b], Cecilia Riccioli[a], Emiliano De Pedro[a], Ana Garrido[a]

[a] Department of Animal Production, E.T.S.I.A.M., Universidad de Córdoba, Campus Rabanales, 14071, Córdoba, Spain

[b] Department of Statistical Science, University College London, 1-19 Torrington Place, WC1E 6BT London, UK

*Corresponding author. Tel.: +34 957 212221; fax: 34 957 218436

E-mail addresses: dcperez@uco.es (D. Pérez-Marín). Animal Production Deparment, ETSIAM, University of Cordoba. Campus Rabanales. 14071. Cordoba, Spain.

Abstract

Iberian pig ham is one of several high value European food products that are the subject of significant attempts at fraud because of the high price differences between commercial categories. Iberian pig products are classified by the Spanish regulations into different categories, mainly depending on the feeding regime during the fattening phase and the race involved, being of Premium quality those products obtained from the animals fed with acorns and other natural resources. Most of the previous NIRS studies related to the Iberian pig have involved the use of at-line instruments to predict quantitative quality parameters. This paper explores the use of the NIR spectra (369 for training and 199 for validation) to classify samples according to the categories Premium (animals fed with acorn) and Non Premium (animals fed with compound feeds), using a MicroNIR™ Pro1700 microspectrometer to analyse individual carcasses *in situ* at the slaughterhouse line. Four discriminant methods were explored: linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), Kernel Bayes and Logistic Regression. These are all discriminant methods that naturally produce classification probabilities to quantify the uncertainty of the results. Rules were tuned and methods compared using both classification error rates and a probability scoring rule. LDA gave the best results, attaining an overall accuracy of 93% and providing well-calibrated classification probabilities.

Keywords: in situ NIRS analysis, portable microspectrometer, Iberian pig classification, carcass authentication, probabilistic discrimination

1. Introduction

The objective of the work reported here was to develop and evaluate classification rules based on NIR *in situ* spectral measurements for the classification of Iberian pig carcasses into Premium (acorn and other natural resources feeding) and Non Premium (compound feeds feeding) categories, with a focus on methods that return probabilities of group membership.

The traditional rearing system of the Iberian pig, which produces the exclusive, delicious and healthy Iberian pig ham, is natural and seems simple. However, in practice, it is highly sophisticated and expensive to undertake. The high prices, ranging from hundreds to thousands of euros, paid for a cured leg of Iberian ham — are an incentive for mislabelling and fraud.

It is well known that the fatty acid content of the adipose tissue and the purity of the breed (percentage of Iberian pig blood) are the main factors affecting the final quality of the Iberian ham [1]. The official quality control systems for determining the feeding regime of the animals are based on on-farm inspection. In addition, sometimes the industry uses the analysis by gas chromatography of the fatty acid composition of samples of melted subcutaneous fat, pooled over animals for reasons of cost. However, these methods are high-cost, time-consuming and only provide information on batches of animals, not on each individual piece of this expensive product. Previous research works have demonstrated that NIRS can be used not only for the prediction of the fatty acid profile, but also that the spectral signal *per-se* can be used to classify Iberian pig carcasses according to the feeding regime during the last fattening phase, and therefore according to commercial categories that command very different prices in the market [2,3,4,5,6].

There is also scientific evidence confirming that NIRS can be used not only *at-line*, but also *on-site*, using handheld NIR sensors - as an IoT (Internet of Things) device - to produce “digital and voluntary labelling systems” to inform producers, industrials, official inspection bodies and consumers about quantitative (fatty acids percentages) and qualitative (feeding

category) attributes of each individual Iberian pig ham [7]. Recently, a new generation of portable, compact and extremely lightweight NIRS instruments has been developed, and these are ideally suited for taking *in situ* measurements. The fast progress in miniaturization of NIR spectrometers has been supported by the development of new micro-technologies such as MEMS (micro-electro-mechanical systems), micro-mirror arrays and LVF (linear variable filters). These technologies have enabled a drastic reduction of spectrometer size and weight while maintaining a good performance [8,9]. These instruments enable the user not only to take spectra at any time, but also to analyse single pieces/units of a food product, and thus obtain more information about the quality and safety of the given product.

In developing an algorithm that converts the NIR spectrum into a classification in an authenticity application, there is a choice between so called one-class modelling, where the authentic class (Premium here) only is modelled and a threshold on this defines the classification, and a more traditional discriminant analysis involving two or more defined classes (Premium and Non Premium here). Oliveri and Downey [10] discuss this choice in the specific context of food authenticity and recommend one-class modelling when the alternative to authentic is ill defined, and discriminant analysis when, as here, there are well defined classes to work with.

In most of the work undertaken for the development of qualitative NIRS classification models, and in most of NIRS food authentication papers, the algorithms used only return a class assignment, i.e. the sample belongs or does not belong to one or more classes. However, for most of the real-word application of NIRS to food authentication issues, and certainly in the case of the Iberian pig ham, it would be extremely valuable to also have information about the uncertainty of the prediction, i.e. to have probabilities associated with class memberships for any unknown predicted sample. There will always be samples for which the assignment is highly reliable and others for which it is relatively uncertain, and the identification of these adds considerable value to the classification.

Probabilistic classifiers have received relatively little attention in NIRS classification work, despite the fact that it will be critical for food authentication, mislabelling or fraud detection issues to use methods that not only provide a prediction or classification result but also information about the uncertainty of the prediction. There are many different algorithms that can be used to build classification rules, too many to try them all on any one application. In deciding which to try here, the choice has been restricted to methods that naturally produce probabilities, as opposed to ones where outputs are, often arbitrarily, transformed to a 0-1 scale and then treated as probabilities. This paper investigates three generative approaches [11] in which probabilistic models are built for the within-class distributions of the spectral data, and one discriminative approach, logistic regression, that directly models the probability of class membership as a function of the spectral data.

2. Material and methods

2.1. Sample sets and reference data

The samples available for training, supplied by 45 different Iberian pig producers, were recorded in two consecutive years (66 samples in 2016 and 429 in 2017). The producers provided the information about the feeding regime of the animals that was used to classify the samples as Premium (animals fed with acorn and other natural resources) and Non Premium (animals fed with compound feeds). In addition, the samples were analysed by Gas Chromatography to determine their fatty acids profile, which is strongly related to the quality of the product and therefore to the feeding regime received by the animals. The final training set comprised 369 of these 495 samples, after those samples with a percentage in oleic acid lower than 52%, but considered by the producers as Premium, were eliminated from the set due to the inconsistency of the reference data. Thus, the distribution of classes in the training set was 139 of Premium and 230 of Non Premium grades.

For the external validation of the models, a new set of 199 samples collected from 10 producers (a subset of the original 45) in the 2018 season was used. The distribution of classes in this set was 105 samples of Premium and 94 of Non Premium grades. No samples were removed from this validation set.

2.2. NIR spectra acquisition

The NIR spectra of subcutaneous adipose tissue of each carcass were collected in the production line at the slaughterhouse, using a microspectrometer MicroNIR™ Pro1700 (VIAVI Solutions, Inc., San Jose, California, USA). The measurements were taken directly in the transversal section of the intact subcutaneous fat beneath the tail insertion area of carcasses, around 2 hours post-slaughter chilled in a temperature controlled chamber (Figure 1).

This ultra compact instrument is based on thin film linear variable filter (LVF) technology for the light dispersing element. It is extremely light and has a window area of around 227 mm^2. The instrument works in reflectance in the spectral region 910 to 1676 nm, taking measurements at 125 spectral points spaced by approximately 6.2 nm. A white reference measurement was obtained using Spectralon™, while a dark reference was obtained from a fixed point in the floor. The sensor integration time was 11 ms and each spectrum was the mean of 200 scans. Each carcass was analysed in duplicate at the slaughterhouse, with one spectrum taken in the inner and other in the outer layer of the subcutaneous fat. Then, before model development, the two spectra were averaged to obtain a single spectrum per sample.

2.3. Classification methodology

Discriminant models were developed to classify the carcasses directly in two quality groups, without going via a quantitative prediction of the fatty acid profile:

- *Premium* category, i.e. those corresponding to animals fed with acorn and other natural resources.
- *NonPremium* category, i.e. those corresponding to animals fed with compound feeds.

In selecting discriminant methods, the focus was on methods that naturally produce classification probabilities, since there will be samples for which the classification is uncertain, and it was considered important to be able to quantify that uncertainty.

Four methods were explored. Three of them, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and kernel Bayes, model the within-class distributions of the spectral data, and use Bayes theorem to pass from probabilities of a spectrum given class membership to probabilities of class membership given the spectrum. With only two classes, it is simplest to use the odds form of Bayes theorem:

$$\frac{P(class = 1|spectrum)}{P(class = 2|spectrum)} = \frac{P(spectrum|class = 1)}{P(spectrum|class = 2)} . \frac{P(class = 1)}{P(class = 2)}$$

Expressed in words, this reads as posterior odds equals likelihood ratio times prior odds. For the results reported below, the prior probabilities of the two classes were taken to be equal, P(class=1) = P(class=2) = 0.5, so that the prior odds become 1. The three methods differ in the way they model the within-class probability distributions of the spectral data, P (spectrum|class=1) and P (spectrum|class=2).

LDA uses multivariate Gaussian distributions with different means but a common covariance matrix, while QDA uses multivariate Gaussian distributions with different means and different covariance matrices, with all these parameters being estimated from the training data [11,12]. The kernel Bayes method builds more flexible models of the within-class distributions by centering a spherical Gaussian distribution (a kernel) on each data point in the training data and then averaging these distributions over each class to construct two probability distributions. This avoids the often unrealistic Gaussian assumption, but at the risk

of over fitting the training data. All three of these approaches are described in much more detail in Fearn et al. [13].

The fourth approach, logistic regression, directly models the probability of class membership as a function of the spectral data. Specifically, the log odds are taken to be linearly related to the vector of spectral data so that:

$$\log_e\left(\frac{P(class = 1|x)}{P(class = 2|x)}\right) = \alpha + \beta^T x$$

where x is the $p \times 1$ vector of spectral data for a sample, α is an intercept and β is a $p \times 1$ vector of coefficients. The coefficients are estimated by maximum likelihood as described in McLachlan [11].

2.4. Spectral pre-treatments

Six spectral pre-treatments were compared. These comprised three levels of derivative treatment, none, first and second, applied to the log (1/R) spectra, without and with the application of the scatter correction SNV (Standard Normal Variate) pre-treatment in each case [14,15]. The first and second derivatives were calculated using a Savitzy-Golay filter, with a second order polynomial and a five-point window [15,16]. SNV was applied after the derivative.

2.5. Dimension reduction

None of the methods described above will work well with 125 highly correlated predictors, so a method of dimension reduction is needed. The obvious choices are to use either principal component analysis (PCA) or partial least squares (PLS) for this. Although it might be argued that PLS scores from a regression on a dummy variable indicating class membership would capture more of the relevant variability than the same number of PCA scores, it was chosen to use PCA. The rationale for this was twofold. Firstly, PCA is simpler and more well defined, in the sense that there is only one set of PC scores for given data whereas PLS scores will be

algorithm dependent, and with plenty of samples there is no problem with using 20 PC scores, which will essentially capture all the relevant variability anyway. Secondly, the PLS factors are extracted to explain covariance, thus concentrating on linear relationships between spectra and class membership. PLS scores might well be a good starting point for LDA, but are not necessarily an improvement over PC scores for the other three methods investigated here. Thus the first q principal component (PC) scores for values of q ranging from 2 to 20 were considered for input to the classification algorithms. The PCA was carried out on the spectra (raw or pre-treated) for the training set only, with spectra for prediction samples being projected into this space. In the case of cross-validation, the PCA was done inside the cross-validation loop.

For the kernel Bayes method, but not for the others, the relative scaling of the PC scores matters. They were scaled to each have standard deviation equal to 1 for the samples used for the PCA.

2.6. Tuning and validation

In the case of LDA, QDA and logistic regression, tuning means selecting a spectral pretreatment and a number of PCs to retain. In the case of the kernel Bayes, there is an additional parameter to tune, the standard deviation (σ) of the kernels. This controls the smoothness of the probability distribution fitted to the spectral data, with larger σ giving smoother distributions. This tuning was carried out using cross validation on the training set. It is generally accepted that leave-one-out-sample cross validation tends to lead to over fitting, and that leaving out blocks is preferable [17].

In this case, there are some natural blocks, as the samples came from 45 producers, and the cross-validation scheme employed was leave-out-one-producer cross validation (LOOPCV). Requiring the classification rule to predict samples from unseen producers makes the cross

validation more like genuine predictions, and could be expected to lead to a more robust choice of parameters.

All four approaches have as their output a probability *p* that the sample whose spectrum was input is of Premium quality. Two assessment criteria were used in the tuning and validation processes. One was to convert the probability into a yes/no classification by assigning the sample to the Premium class if $p > 0.5$ and to count the number of misclassifications. The other was to assess the probabilities themselves by calculating a log probability score:

$$L = \frac{1}{n}\sum_{i=1}^{n} \log_e(p_i),$$

where the sum is over the *n* samples being predicted and p_i is the probability assigned by the classification rule to the true class of the *i*th sample. The best possible score would be zero, achieved by giving probability 1 to the true class in each case. Giving probability 0.5 of being Premium to all samples would give a score of $\log_e(0.5) = -0.69$, so any useful rule should have a score in the range from -0.69 to 0. This log score heavily penalizes a rule for being over confident. For example, assigning a probability of 0.999 to the class to which the sample does not belong will score $\log_e(0.001) = -6.9$ for that sample.

2.7. Computing

The computations were programmed in Matlab software version 2016b (The MathWorks, Inc., Natick, MA, USA). The logistic regression was implemented using the Statistics and Machine Learning Toolbox, and the pretreatments via the PLS Toolbox version 8.7 (Eigenvector Research, Manson WA, USA).

3. Results

3.1. Exploration of the data

Figure 2 shows the SNV treated second derivative spectra for the training set of 369 samples, from 2016 and 2017. The Premium samples are plotted in grey, and the NonPremium samples

in black. The NonPremium samples were plotted second, and cover many of the Premium ones, but there are regions, especially that between 1400 and 1450nm, where there appears to be some separation. Several authors have demonstrated that the region 1436-1442 nm, among others, is characteristic of CH_2 absorptions related to fatty acid profile of vegetable and animal fats and oils [18]. It is well known that the main differences between the Iberian products coming from animals with different feeding regime (acorn *vs* compound feeds) are related to the fatty acid profile of the fat [1,5,19].

Figure 3 shows the scores on the first two PCs from a PCA of these 369 samples, with the Premium and NonPremium classes distinguished by plotting symbol and colour. Although there is overlap between the classes, there is enough separation to suggest that classification may be possible using these spectra, especially if more PCs are used. The other obvious comment here is that the with-class distributions do not look Gaussian, which may favour kernel Bayes over the methods that assume Gaussian distributions.

3.2. LDA model

The LDA model was trained on the 2016 and 2017 samples using leave-out-one-producer cross validation (LOOPCV). Six pretreatments, as described in the methodology section, were tried, as were numbers of principal components from 2 to 20. Table 1 shows the optimal number of PCs for each pretreatment, chosen to optimise either the number of incorrect classifications or the probability score. The training set has 369 samples, so the numbers of errors in the third column of the Table 1 correspond to error rates ranging from 7.6% (28/369) to 9.2% (34/369). Not surprisingly, the two criteria lead to different choices of optimal pretreatment and number of PCs. To look for good compromise choices, the plots in Figure 4 were inspected. To get the results for both criteria onto comparable scales, they were expressed as % worse than the global best scores, which are 28 errors for raw spectra using 11 PCs in the LDA and a probability score of −0.228 for second derivative spectra using 4 PCs.

There are several possible candidates for a compromise choice. Second derivative plus SNV with 8 PCs, which gives 29 errors (7.9%) and a probability score of –0.240, was selected, partly because the performance seems fairly stable around this choice of number of PCs. For the main competitor, raw spectra with 13 PCs, one PC either more or less leads to a significant deterioration in performance on one or other criterion.

The selected model was then used to predict the 2018 samples. It made 13 errors out of 199 classifications, an error rate of 6.5%. Of the 13 errors, 2 involved the classification of Premium samples as Non Premium; 11 were in the other direction.

To assess the calibration of the probabilities assigned by the LDA rule, the 199 predicted samples were classified into 10 bins according to the probability assigned to the Premium category. The bins are of unequal width, partly because the probability scale is intrinsically non linear, partly in order to avoid bins with very small numbers of samples. Table 2 shows the number of samples and the number and proportion of Premium samples in each bin. The results look quite reasonable, with 9 of the 13 errors occurring in the probability range 0.2 – 0.8, and no errors in the bins where probabilities are less than 0.01 or greater than 0.99. The number of extreme probabilities near 1 is possibly a cause for concern. It is impossible to assess their accuracy without thousands of samples but this level of certainty, arising as it does from the use of light tailed Gaussian distributions in the model, may well be over confident.

3.3. QDA model

Training QDA in the same way as LDA gave the results shown in Table 3.

Examination of plots like those in Figure 2 led to the selection of “second derivative + SNV” with 3 PCs as the best model. This gave 30 errors (8.1%) and a probability score of –0.326. Applied to the validation set of 199 samples, the QDA rule made 41 errors, with 15 Premium samples classified as Non Premium and 26 Non Premium samples classified as Premium.

This is an error rate of 21%. The results in Table 4 suggest that although the probabilities in the midrange seem reasonable, the more extreme ones, especially as they approach 1, are over confident.

3.4. Kernel Bayes

Training this method involves optimising the pretreatment and the number of PCs used as for the other methods, but also the standard deviation σ of the kernels. Table 5 shows the optimal choices via cross validation of the number of PCs and σ for each pretreatment, again using both number of errors and the probability score as criteria to optimise.

The model using "second derivative + SNV" pretreatment with 4 PCs and $\sigma = 0.7$ was chosen and used to predict the 2018 validation samples. This gave 39 errors, of which 13 were Premium classified as Non Premium and 26 Non Premium classified as Premium. This is an error rate of 20%, much higher than the 25/369=6.8% on the training set. The performance of the classification probabilities is shown in Table 6. This method gives much more conservative probabilities than either QDA or LDA. There are no errors for probabilities less than 0.01 or greater than 0.99, but only 20 of the 199 probabilities are in this range.

3.5. Logistic regression model

Training Logistic Regression in the same way as the other methods gave the results shown in Table 7. From the two obvious candidates (see Table 7), "second derivative + SNV" with 4 PCs was chosen, because of its optimality for the other methods.

Applied to the validation set there were 35 errors among the 199 samples, 23 of these being Premium samples misclassified as Non Premium and 12 being Non Premium misclassified as Premium. This is an 18% error rate, compared with only 7.9% on the training set. The performance of the probabilities, as shown in Table 8, is not unreasonable: there is only one error in the 63 samples with probabilities less than 0.01 or greater than 0.99 for example.

3.6. Comparison of the 4 methods and discussion of the results

Table 9 shows two measures of performance on the validation set for each of the 4 methods employed: the number of classification errors and the probability score.

LDA is a clear winner on both criteria, with its relatively low probability score reflecting the fact that not only does it make fewer mistakes, it assigns moderate probabilities to the misclassified samples as can be seen from Table 2. That LDA beats the more sophisticated QDA and kernel Bayes on the validation set is probably due to over fitting on the part of the latter two. Simpler methods like LDA tend to extrapolate better to new situations, in this case to predictions for a year not seen. It is less obvious why the logistic regression, which is also relatively simple, fails to predict well.

To produce a classifier for future samples, the obvious approach is to fix the method as LDA with 8 PC's and apply it to the combined data for the three years. Doing this and assessing by LOOPCV gave 42 errors out of 568 classifications, equal in number to the 29 cross validation errors plus 13 validation errors reported above, for an overall error rate of 7.3%, i.e. a total percentage of correctly classified samples of 92.7%.

The LDA model can be expressed in terms of a linear discriminant function (Figure 5). It can be appreciated that the main peaks at 1160, 1190 and 1380 nm are regions related with the absorption of CH_2 bonds [20]. As has been commented earlier, this group is related with fat and the fatty acid profile of the product, this being one of the main effects of the different feeding regimes on the quality of Iberian pork products [1,19].

In a previous study, Zamora et al. [6] also analysed *in situ* Iberian pig carcasses, using in this case a MEMS handheld instrument and discriminating between three categories according to the old Spanish regulations. They reported percentages of total correctly classified samples of 83.6 % in validation using PLS2 discriminant analysis, considerably lower than the 92.7% achieved in this work.

No previous studies of this type have reported the uncertainty of the prediction in the classification models, as is done here. This is a key point for the implementation of NIR sensors in the industry for classification purposes. With any classification based on a rapid measurement, there will be some samples where the assignment is quite uncertain. If it is known which samples these are, they can be investigated further using more accurate (and more expensive) tests if the context merits it. Among the 169 validation samples assigned probabilities of less than 0.05 or greater than 0.95 of being Premium there are only 2 errors. When the rapid test says it is confident, it can be relied on.

The overall error rate of 7% needs to be considered in the light of the probabilities. The majority of the errors (9/13) occur in the 15% of the validation samples assigned probabilities between 0.2 and 0.8. Of course it would be desirable to reduce the proportion of samples for which there is this much uncertainty about the classification. This might be achieved through improvements to the instrumentation and the measurement procedures as well as to the classification algorithms. However it should be remembered that the natural variation between animals and the possibility of feeding mixed diets will mean that there will always be borderline cases.

The probabilities have been converted into classifications using the obvious cutoff of 0.5. For the validation set, this has resulted in 2 false negatives and 11 false positives, an unfortunate split in the context of consumer protection. Moving the threshold to 0.8 instead would give 9 false negatives and only 4 false positives, i.e. 4 Non Premium samples classified as Premium. Such a threshold could be formally justified by introducing asymmetric classification error losses into the calculation. To do this with the current data risks over interpreting small numbers of errors, but it might be appropriate were many more validation samples to become available.

4. Conclusions

The portable NIR microspectrometer device enables the *in-situ* inspection of individual Iberian pig carcasses at the speed of the slaughtering process. The best classification model developed, based on the use of LDA, has shown the possibilities of this technology for classifying the carcasses according to the categories Premium and Non Premium, related to the feeding regime of the animals during the fattening phase (acorn *vs* compound feeds). The LDA model achieved an overall classification accuracy of 93% and was able to output classification probabilities that appeared to be well calibrated. The potential of this kind of application, able to give predictions with associated probabilistic uncertainty, opens new and enormous possibilities for the use of NIRS for food integrity and authenticity issues.

Acknowledgments

This research was framed within one of the world's largest research projects in the area of food authenticity and traceability: the FoodIntegrity project, EU sponsored (7th FP).

References

[1] A. Garrido-Varo, E. De Pedro, The role of Near-infrared Spectroscopy in verifying label information in agro-forestry, in: D.A. Burns, E.W. Ciurczak (Eds.), Handbook of Near-Infrared Spectroscopy Analysis, Third Edition, Practical spectroscopy series volume 35, CRC Press, USA, 2007, pp. 387-398.

[2] C. Hervás, A. Garrido, B. Lucena, N. García, E. De Pedro, Near infrared spectroscopy for classification of Iberian pig carcasses using an artificial neural network, Journal of Near Infrared Spectroscopy, 2 (1994) 177-184. https://doi.org/10.1255%2Fjnirs.44.

[3] E. De Pedro, A. Garrido, A. Lobo, P. Dardenne, I. Murray, Objective classification of Iberian pig carcasses: GC versus NIR, in: G. D. Batten, P. C. Flinn, L. A. Welsh, A. B. Blakeney (Eds.), Leaping ahead with near infrared spectroscopy, NIR Spectroscopy Group, Royal Australian Chemistry Institute, Melbourne, Australia, 1995, pp. 291–295.

[4] J. García-Olmo, A. Garrido, E. De Pedro, The transfer of fatty acid calibration equations using four sets of unsealed liquid standardisation samples, Journal of Near Infrared Spectroscopy, 9 (2001) 49-62. https://doi.org/10.1255%2Fjnirs.294.

[5] D. Pérez-Marín, E. De Pedro, J.E. Guerrero, A. Garrido-Varo, A feasibility study on the use of near-infrared spectroscopy for prediction of the fatty acid profile in live Iberian pigs and carcasses, Meat Science 83 (2009) 627-633.
https://doi.org/10.1016/j.meatsci.2009.07.012

[6] E. Zamora-Rojas, D. Pérez-Marín, E. De Pedro-Sanz, J.E. Guerrero-Ginel, A. Garrido-Varo, In-situ Iberian pig carcass classification using a micro-electro-mechanical system (MEMS)-based near infrared (NIR) spectrometer, Meat Science, 90 (2012) 636-642. https://doi.org/10.1016/j.meatsci.2011.10.006.

[7] A. Garrido-Varo, C. Riccioli, T. Fearn, E. De Pedro, D. Pérez-Marín, Multivariate predictive models for the prediction of fatty acids in the EU high added-value "acorn Iberian pig ham" using a miniature near-infrared spectroscopy instrument, Proceedings of SPIE 11016, Sensing for Agriculture and Food Quality and Safety XI, 110160C, 2019. https://doi.org/10.1117/12.2521977.

[8] R.A. Crocombe, Portable Spectroscopy, Applied Spectroscopy, 72 (2018) 1701–1751. https://doi.org/10.1177%2F0003702818809719.

[9] H. Yan, H.W. Siesler, Hand-held near-infrared spectrometers: state-of-the-art instrumentation and practical applications, NIR News 29 (2018) 8–12. https://doi.org/10.1177%2F0960336018796391.

[10] P. Oliveri, G. Downey, Multivariate class modelling for the verification of food-authenticity claims. Trends in Analytical Chemistry, 35 (2012) 74-86. https://doi.org/10.1016/j.trac.2012.02.005.

[11] G.J. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, Chapter 7, Wiley, New York, 1992.

[12] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning. Springer, New York, 2001.

[13] T. Fearn, D. Pérez-Marín, A. Garrido-Varo, J.E. Guerrero-Ginel, Classifying with confidence using Bayes rule and kernel density estimation, Chemometrics and Intelligent Laboratory Systems, 189 (2019) 81-87. https://doi.org/10.1016/j.chemolab.2019.04.004.

[14] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near infrared diffuse reflectance spectra, Applied Spectroscopy 43 (1989) 772–777. https://doi.org/10.1366/0003702894202201.

[15] T. Naes, T. Isaksson, T. Fearn, T. Davies, A user Friendly guide to Multivariate Calibration and Classification, NIR Publications, Chichester, UK, 2002.

[16] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, Analytical Chemistry, 36 (1964) 1627–1639. https://doi/abs/10.1021/ac60214a047.

[17] B. D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, UK, 1996.

[18] A. Garrido, J. García, D. Pérez, Applications in Fats and Oils, in: C. Roberts, J. Workman, J. Reeves (Eds.), Near-Infrared Spectroscopy in Agriculture, Chapter 19, Agronomy Monograph, 44, USA, 2004, pp. 487-558.

[19] S. Ventanas, J. Ventanas, J. Ruíz, M. Estévez, Iberian pigs for the development of high-quality cured products, Recent Res. Devel. Agricultural & Food Chem., 6 (2005) 1-27.

[20] J.S. Shenk, J.Jr. Workman, M.O. Westerhaus, Application of NIR spectroscopy to agricultural products, in: D.A. Burns, E. Ciurczak (Eds), Handbook of Near-Infrared Analysis, Third Edition, Practical spectroscopy series volume 35, CRC Press, USA, 2007, pp. 347–386.

Table 1. Optimal numbers of PCs chosen for LDA, either to minimise the number of errors or to maximise the probability score, using cross validation on the 2016 and 2017 samples.

Pre-treatment	**Minimise no. of errors**		**Maximise prob. score**	
	PCs	**Errors**	**PCs**	**P score**
Raw	11	28	13	–0.237
SNV	2	31	16	–0.251
d1	6	31	12	–0.230
d1+SNV	19	34	13	–0.254
d2	9	30	4	–0.228
d2+SNV	3 or 8	29	7	–0.236

PCs: Principal Components; d1: first derivative; d2: second derivative;
SNV: Standard Normal Variate

Table 2. Proportion of true Premium samples in each of 10 ranges for the probability assigned to Premium by the LDA rule.

Prob. Premium	No. samples	No. Premium	Prop. Premium
0 – 0.001	5	0	0.00
0.001 – 0.01	14	0	0.00
0.01 – 0.05	28	0	0.00
0.05 – 0.2	23	0	0.00
0.2 – 0.5	15	2	0.13
0.5 – 0.8	15	8	0.53
0.8 – 0.95	16	14	0.88
0.95 – 0.99	27	25	0.93
0.99 – 0.999	27	27	1.00
0.999 – 1	29	29	1.00

Table 3. Optimal numbers of PCs chosen for QDA, either to minimise the number of errors or to maximise the probability score, using cross validation on the 2016 and 2017 samples.

Pre-treatment	**Minimise no. of errors**		**Maximise prob. score**	
	PCs	**Errors**	**PCs**	**P score**
Raw	4	38	2	–0.366
SNV	2	34	3	–0.306
d1	10	40	2	–0.424
d1+SNV	5	39	2	–0.397
d2	14	37	3	–0.301
d2+SNV	3	30	4	–0.267

PCs: Principal Components; d1: first derivative; d2: second derivative;
SNV: Standard Normal Variate

Table 4. Proportion of true Premium samples in each of 10 ranges for the probability assigned to premium by the QDA rule.

Prob. Premium	No. samples	No. Premium	Prop. Premium
0 – 0.001	6	0	0.00
0.001 – 0.01	8	0	0.00
0.01 – 0.05	21	3	0.14
0.05 – 0.2	23	3	0.13
0.2 – 0.5	25	9	0.36
0.5 – 0.8	22	14	0.64
0.8 – 0.95	20	11	0.55
0.95 – 0.99	25	20	0.80
0.99 – 0.999	18	14	0.78
0.999 – 1	31	31	1.00

Table 5. Optimal numbers of PCs and kernel standard deviation σ chosen for kernel Bayes, either to minimise the number of errors or to maximise the probability score, using cross validation on the 2016 and 2017 samples.

Pre-treatment	**Minimise no. of errors**			**Maximise prob. score**		
	PCs	**σ**	**Errors**	**PCs**	**σ**	**P score**
Raw	2	0.6	30	2	0.5	–0.358
SNV	4	0.7	33	4	0.7	–0.312
d1	6	0.9	43	7	0.9	–0.379
d1+SNV	4	1.3	31	4	0.6	–0.301
d2	4	0.5	28	4	0.6	–0.266
d2+SNV	4	0.7	25	4	0.7	–0.282

PCs: Principal Components; d1: first derivative; d2: second derivative;
SNV: Standard Normal Variate

Table 6. Proportion of true Premium samples in each of 10 ranges for the probability assigned to Premium by the kernel Bayes rule.

Prob. Premium	No. samples	No. Premium	Prop. Premium
0 – 0.001	0	0	-
0.001 – 0.01	1	0	0.00
0.01 – 0.05	2	0	0.00
0.05 – 0.2	28	1	0.04
0.2 – 0.5	50	12	0.24
0.5 – 0.8	45	30	0.67
0.8 – 0.95	41	33	0.80
0.95 – 0.99	13	10	0.77
0.99 – 0.999	12	12	1.00
0.999 – 1	7	7	1.00

Table 7. Optimal numbers of PCs chosen for Logistic Regression, either to minimise the number of errors or to maximise the probability score, using cross validation on the 2016 and 2017 samples.

Pre-treatment	**Minimise no. of errors**		**Maximise prob. score**	
	PCs	**Errors**	**PCs**	**P score**
Raw	13	33	11	–0.320
SNV	2	33	6	–0.348
d1	8	31	8	–0.257
d1+SNV	13	33	5	–0.336
d2	4	29	4	–0.243
d2+SNV	4	29	4	–0.247

PCs: Principal Components; d1: first derivative; d2: second derivative;
SNV: Standard Normal Variate

Table 8. Proportion of true Premium samples in each of 10 ranges for the probability assigned to Premium by the Logistic Regression rule

Prob. Premium	No. samples	No. Premium	Prop. Premium
0 – 0.001	5	0	0.00
0.001 – 0.01	24	1	0.04
0.01 – 0.05	22	4	0.18
0.05 – 0.2	20	7	0.35
0.2 – 0.5	20	11	0.55
0.5 – 0.8	19	13	0.68
0.8 – 0.95	20	16	0.80
0.95 – 0.99	21	19	0.90
0.99 – 0.999	8	8	1.00
0.999 – 1	26	26	1.00

Table 9. Performance of four methods on the 199 validation samples compared on two criteria, the number of errors and the probability score.

	No. of PCs	No. errors	Prob. score
LDA	8	13	–0.176
QDA	3	41	–0.548
Kernel Bayes	4	39	–0.462
Logistic Regression	4	35	–0.407

PCs: Principal Components

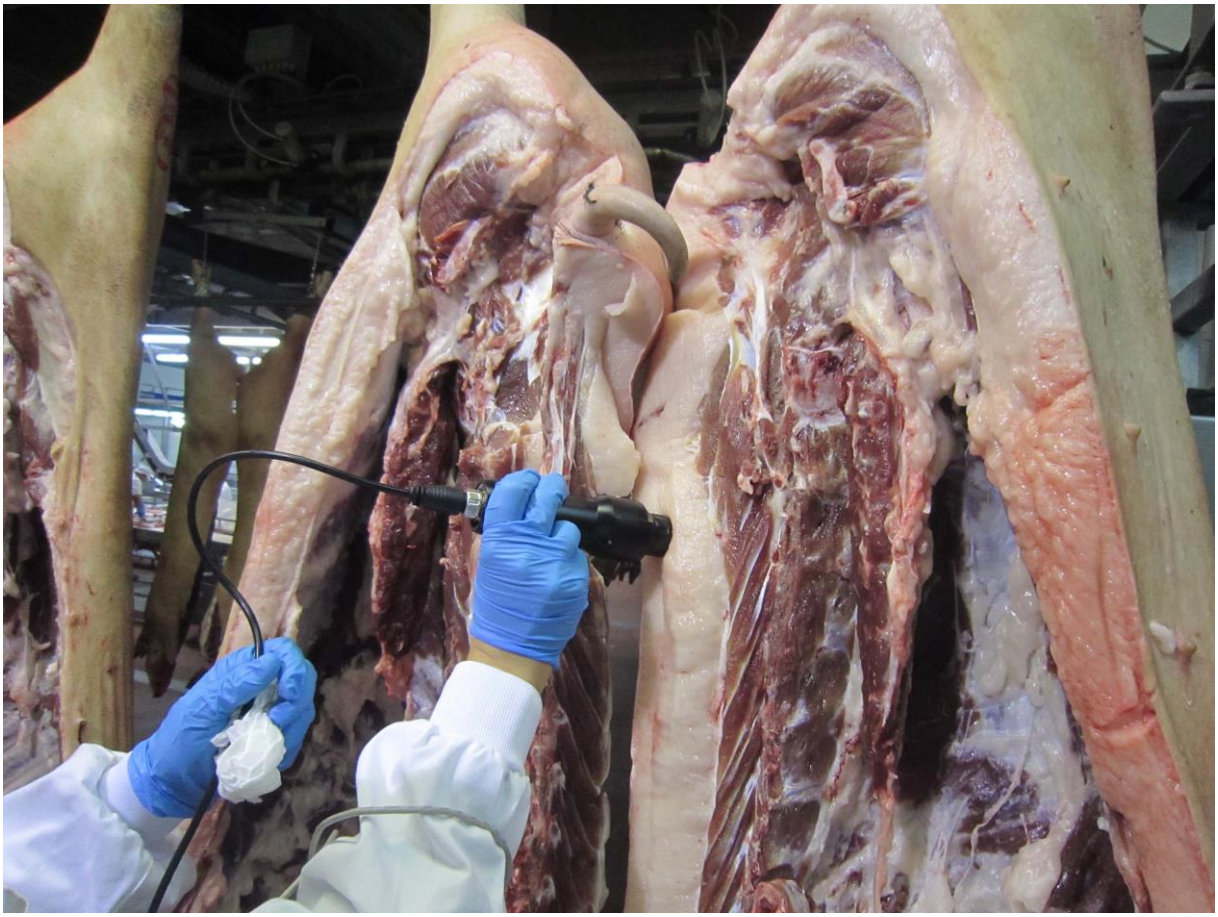

Figure 1. In-situ NIRS carcass analysis.

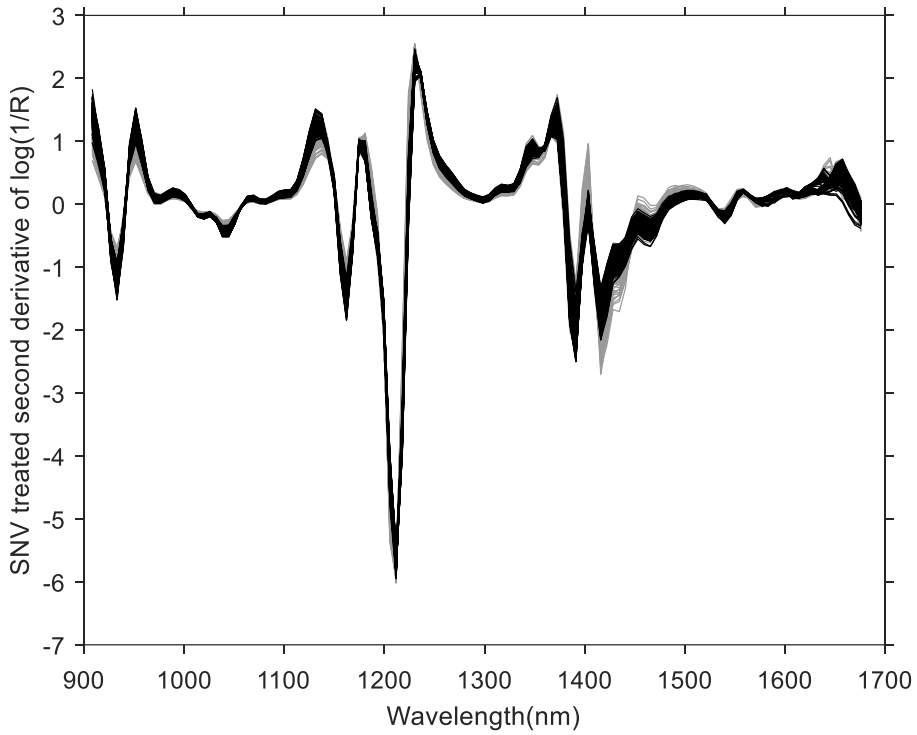

Figure 2. SNV treated second derivative spectra for the 369 samples in the training set (2016 and 2017). Grey = Premium, black = Non Premium.

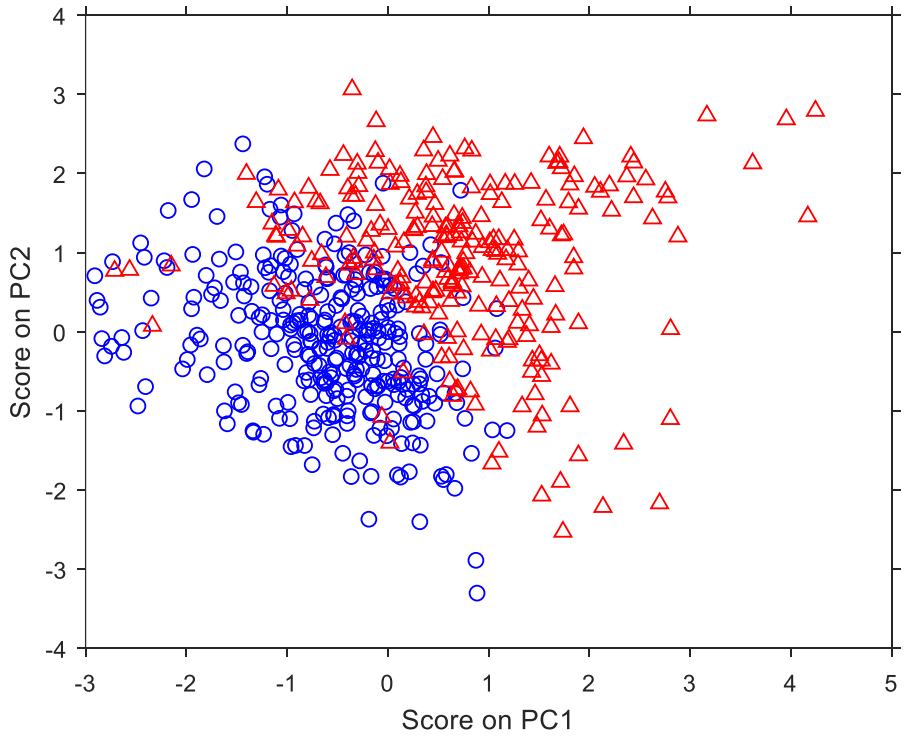

Figure 3. Scores on the first two PCs from a PCA of the SNV treated second derivative spectra of the 369 samples in the training set (2016 and 2017). Red triangle = Premium, blue circle = Non Premium.

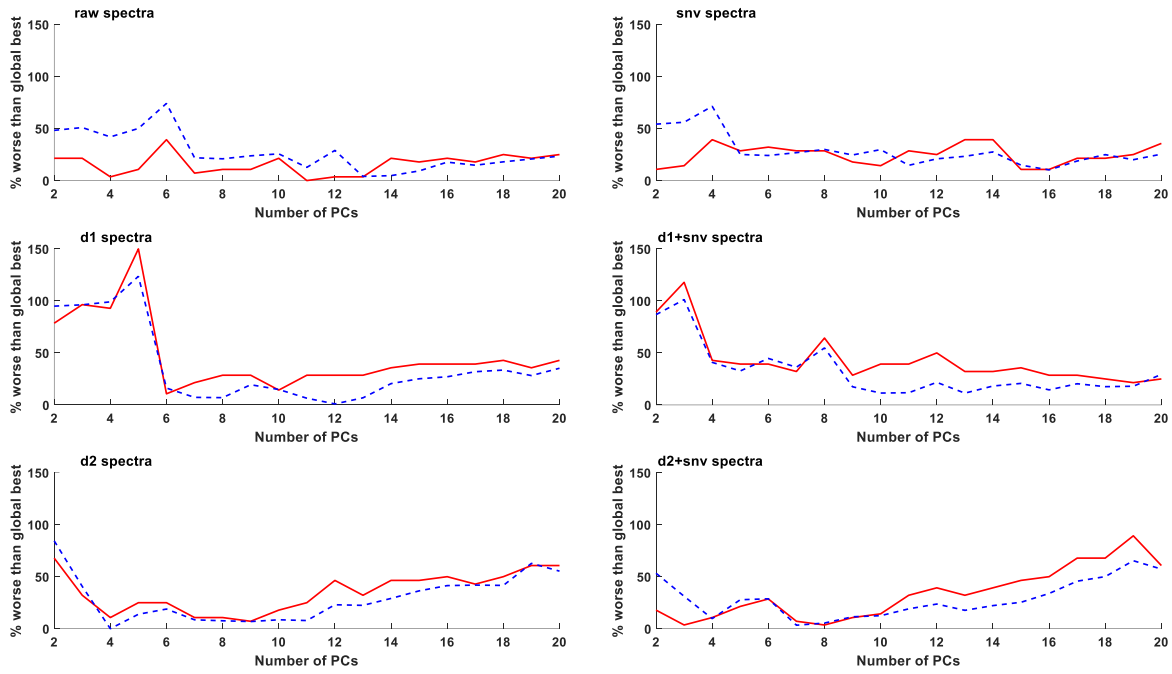

Figure 4. Scores on two criteria for 6 pretreatments as a function of numbers of PCs used in LDA. Solid red line = number of errors, dashed blue line = probability score, both expressed as % worse than the global best results of 28 errors (raw, 11PCs) and –0.228 (d2, 4PCs).

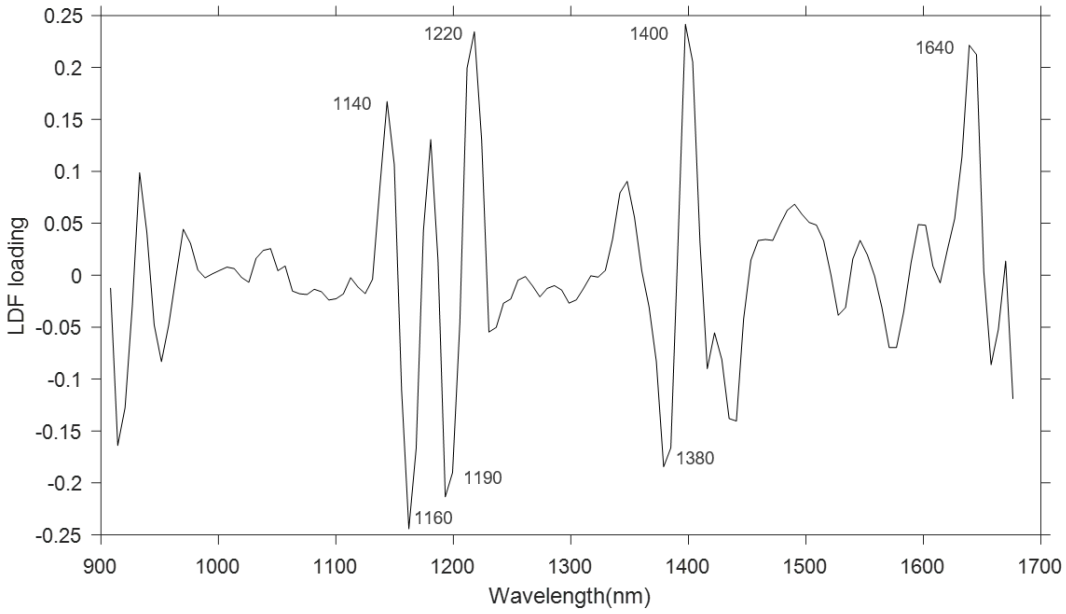

Figure 5. Loading vector for linear discriminant function derived from LDA on 2016+2017 samples using 8 PCs calculated from second derivative + SNV pretreated spectra.