Journal of Food Engineering 90 (2009) 540–547

Contents lists available at ScienceDirect

Journal of Food Engineering

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

Duroc and Iberian pork neural network classification by visible and near infrared reflectance spectroscopy

F.G. del Moral [a], A. Guillén [b], L.G. del Moral [c], F. O'Valle [a], L. Martínez [d], R.G. del Moral [a,*]

[a] Department of Pathology, School of Medicine, University of Granada, Avda. de Madrid, 11, 18012 Granada, Spain
[b] Department of Computer Technology and Architecture, ETSI Computer Science, University of Granada, Spain
[c] Department of Vegetal Physiology, School of Science, University of Granada, Spain
[d] Department of Chemical Engineering, School of Science, University of Granada, Spain

ARTICLE INFO

Article history:
Received 29 January 2008
Received in revised form 3 June 2008
Accepted 21 July 2008
Available online 3 August 2008

Keywords:
Pork
Pig
Iberian
Duroc
Mutual information
Neural network
Spectra characterization
VIS/NIRS

ABSTRACT

Visible and near infrared reflectance spectroscopy (VIS/NIRS) was used to differentiate between Duroc and Iberian pork in the M. *masseter*. Samples of Duroc (n = 15) and Iberian (n = 15) pig muscles were scanned in the VIS/NIR region (350–2500 nm) using a portable spectral radiometer. Both mutual information and VIS/NIRS spectra characterization were developed to generate a ranking of variables and the data were then processed by artificial neural networks, establishing 1, 3, or 10 wavelengths as input variable for classifying between the pig breeds. The models correctly classified >70% of all problem assumptions, with a correct classification of >95% for the three-variable assumption using either mutual information ranking or VIS/NIRS spectra characterization. These results demonstrate the potential value of the VIS/NIRS technique as an objective and rapid method for the authentication and identification of Duroc and Iberian pork.

1. Introduction

Iberian pig (*Sus mediterraneus*) is a rustic breed from the Southeast of the Iberian Peninsula with a high adipogenic potential. These animals are reared naturally in *La Dehesa*, a natural ecosystem largely comprising cork trees, holm oaks, and gall-oaks, and undergo a final fattening phase based on the consumption of natural resources, mainly acorns, in extensive acorn-rich pasture lands, called *montanera* in Spanish. In the past, the economic yield from this breed was low, but a recent high demand for Iberian dry-cured products led to an increase in the Iberian pig census to one million pigs in 2004 (MAPA, 2004). The Iberian breed has traditionally been crossed with Duroc pigs to obtain more piglets per sow and a higher weight at weaning and at the end of fattening (Ramírez and Cava, 2007). Crossbred Iberian–Duroc pigs have more lean meat, less fat, and higher growth rates versus pure Iberian pigs. Since 2001, Spanish legislation (BOE, 2001) has only permitted Iberian × Duroc crosses when the maternal line is pure Iberian, in order to preserve the purity of the Iberian breed.

Ramírez and Cava (2007) concluded that the crossing of Duroc genotypes selected for meat production with Iberian pigs significantly reduced the quality of Iberian products, resulting in less marbling and lower intramuscular fat contents, less intensity of color and odor, poor texture characteristics, higher polyunsaturated fatty acids content, more susceptibility to rancidity, and a saltier taste. They proposed that Duroc lines crossed with Iberian pigs required characterization to avoid losses in the quality of Iberian products.

With respect to characterization, visible and near infrared spectroscopy (VIS/NIRS) has become established over the past 30 yrs as a rapid and effective method to assess the quality of meat products. A major challenge faced by the industry is the determination of the authenticity and purity of its products with the possibility of integrating this assessment into an analytical system of surveillance for certified meat products by rapid screening techniques (Hoving-Bolink et al., 2005; Savenije et al., 2006). Consumers demand accurate and fast quality control in the different phases of the manufacturing process, and VIS/NIRS has emerged as an appropriate tool for quality assurance because optical methods are non-destructive, fast, inexpensive, portable, and are considered suitable for online measurements, providing both consumers and suppliers with information about the product they are buying or selling (Swatland, 1994; Xia et al., 2007).

According to the literature, VIS/NIRS technology has been used in pork to determine intramuscular fat (Hoving-Bolink et al., 2005; Savenije et al., 2006), fatty acid composition (Fernández-Cabanás

* Corresponding author. Tel.: +34 958244097; fax: +34 958243510.
E-mail address: rgarcia@ugr.es (R.G. del Moral).

0260-8774/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jfoodeng.2008.07.027

et al., 2007; González-Martín et al., 2005; González-Martín et al., 2003), color (Cozzolino et al., 2003), water-holding capacity (Brøndum et al., 2000), and presence of RN^- genetic allele (Josell et al., 2000), but it has not been applied for the direct qualitative classification of meats of varied quality and price (i.e., gastronomic value). Novel methodologies are required for the traceability of meats, establishing their authenticity and detecting adulteration (Cozzolino and Murray, 2004; Ortiz-Somovilla et al., 2005; Ortiz-Somovilla et al., 2007).

The objective of this study was to examine the feasibility of VIS/NIRS to differentiate Iberian pork from standard Duroc pork in fresh meat and to evaluate the accuracy of prediction models, comparing neural network based on mutual information ranking with classical models based on the VIS/NIRS spectrum. This paper proposes a synergy of NIRS and neural networks for the automatic classification of Iberian × Iberian and Duroc × Iberian breeds.

2. Material and methods

2.1. Experimental design

Thirty animals were slaughtered at a commercial slaughter plant: one group of 15 Duroc pigs given standard feed and another of 15 Iberian pigs that fed in the open on acorns and pasture (''*montanera*''). At slaughter, the Duroc pigs had mean age of 6 months and mean carcass weight of 85 kg, whereas the Iberian pigs had mean age of 12–14 months and mean carcass weight of 125 kg. After exsanguination, 2 h hygienic-health analysis of carcasses was performed in a refrigeration chamber a 4 °C, followed by hot deboning and extirpation of the right-hand side of M. *masseter*. The muscle was vacuum-packed and preserved in refrigerated chamber at 4 °C for 72 hrs, thereby simulating the mean transport time from the slaughterhouse to the supermarkets for sale to the final consumer. Three hours before quantifications were performed, the vacuum bag was opened to allow oxygenation of the meat and avoid interferences due to oxidation state of the myoglobin.

2.1.1. Macroscopic description of M. masseter

M. *masseter* presents some specific histological characteristics, with a major development of intramuscular connective tissue (IMCT), especially of the perimysium, due to the high tension it must endure. The outer fascia is lean, with little surface marbling. Its macroscopic histology is highly homogeneous, with a large predominance of striated muscle fibers. In contrast, the inner fascia delimiting the oral cavity shows major marbling with macroscopically visible strips of IMCT several millimeters wide. For consistency, the outer side of the muscle is designated the lean side and the inner side the marbled side.

2.2. VIS/NIRS measurements

2.2.1. Spectrometer

The reflectance spectrum of M. *masseter* was acquired using a portable spectral radiometer (FieldSpec® Pro JR A 110080; Analytical Spectral Devices Inc., Boulder, CO) with a reflectance range of 350–2500 nm. The sampling interval was 1.4 nm for the spectral range 350–1000 nm and 2 nm for the range 1000–2500 nm. The system also incorporated a contact probe (FieldSpec Pro FR A111208; Analytical Spectral Devices Inc.) with halogenous light source and measurement surface area equivalent to a circle 2 cm in diameter, and a maximum specular reflectance of 5%.

2.2.2. Spectra

White level was calibrated on a Spectralon of 3.62″ diameter (Analytical Spectral Devices Inc.). Material with approximately 100% reflectance across the entire spectrum is designated a white reference panel or white reference standard.

The spectrum was optimized by adjusting sensitivity to varying conditions of illumination with software-adjustable integration time, which determines how long the array will collect energy in the 350–1050 nm region, and uses automatically adjusted gains in the 1000–2500 nm regions to amplify the signal. A dark current correction (DCC) was applied to remove electrical current generated by thermal electrons (called dark current), and is added to that generated by incoming photons. The raw data returned are 16-bit numbers corresponding to the output of each element in the VIS/NIRS detector array and each 2 nm sample of the spectrum to generate a relative reflectance (the reflectance factor), which is the quantity that the system measures. The system reduces noise in the desired spectral signal by spectrum averaging technique (average of 20 spectrums per quantification).

Two spectra were acquired for each animal at 72 h post-mortem, directly applying the contact probe on both the lean side and marbled side of the M. *masseter*.

2.3. Data analysis

The data analysis had two main objectives: 1) to establish a ranking of wavelengths, using mutual information ranking (MI), to identify relevant areas of the spectra for the data classification; 2) to use this ranking to classify the data by means of radial basis function neural networks (RBFNNs).

2.3.1. Mutual information ranking

Mutual information (MI) theory (cross-entropy) was applied to determine the most representative wavelengths in the differentiation between Duroc and Iberian pork and to characterize the spectra. Given the two variables $X = \{x_k, k = 1 \dots n\}$ and $Y = \{y_k, k = 1 \dots n\}$, the mutual information between X and Y can be defined as the amount of information that the variables X provide about Y and vice-versa, and can be expressed as:

$$I(X|Y) = H(Y) - H(Y/X)$$

where $H(Y)$ is the entropy of variable Y, which measures the uncertainty on Y. In the continuous case and according to the formulation of Shannon, it is defined as:

$$H(Y) = -\int \mu_Y(y)\log\mu_Y(y)\mathrm{d}y,$$

where $\mu_Y(y)$ is the marginal density function, which can be defined using the joint probability density function (PDF) $\mu X, Y$ of X and Y as

$$\mu_Y(y) = \int \mu_{X,Y}(x,y)\mathrm{d}x,$$

and $H(Y/X)$ is the conditional entropy, which measures the uncertainty of Y, since X is known. It is defined in the continuous case as:

$$-\int \mu_X(x)\int \mu_Y(y|X=x)\log\mu_Y(y|X=)\mathrm{d}y\mathrm{d}x.$$

In other words, mutual information $I(X,Y)$ is the decrease in uncertainty on Y once X is known. Due to entropy properties, mutual information can also be defined as:

$$I(X,Y) = H(X) + H(Y) - H(H|Y),$$

leading to:

$$I(X,Y) = \int \mu_{X,Y}(x,y)\log\frac{\mu_{X,Y}(x,y)}{\mu_X(x)\mu_Y(y)}\mathrm{d}x\mathrm{d}y.$$

Hence, only the estimate of the joint PDF between X and Y is needed to estimate mutual information between two groups of variables.

Estimating the joint probability distribution can be performed using a number of techniques (Bonnlander and Weigend, 2004; Herrera et al., 2006). The estimator used in the present study is based on k-nearest neighbors (k-nn) because it has some advantages over estimators based on histograms or kernels and has been shown to perform adequately (Kraskov et al., 2004). This MI estimator (Astakhov et al., 2007) is dependent on the value chosen for k (k-th nearest neighbor). As recommended by Stögbauer et al. (2004), A is used for a trade-off between variance and bias, with a mid-range value for k (k = 6).

2.4. Radial basis function neural networks (RBFNN)

An RBFNN is a two-layer, fully connected network in which each neuron implements a Gaussian function as follows:

$$e^{-\frac{\|\vec{x}_k-\vec{c}_i\|^2}{r_i^2}},$$

where $\vec{x}_k$ is an input vector, $\vec{c}_i$ is a center of the ith RBF and r_i^2 is the radius of the ith RBF.

The output layer implements a weighted sum of all the outputs from the hidden layer:

$$F(\vec{x}_k; C, R, \Omega) = \sum_{i=1}^{m} \Phi(\vec{x}_k; \vec{c}_i r_i) \cdot w_i,$$

where $\Omega = \{w_i, \ldots, w_m\}$ are the output weights, which modulate the contribution of a hidden layer to the corresponding output unit and can be obtained optimally by solving a linear equation system. The network architecture is depicted in Fig. 1. Once the continuous output is obtained, the class corresponding to the input sample is computed by rounding the output of the function to its closest integer, i.e., 0 for Duroc pork and 1 for Iberian pork.

This kind of neural network has been applied successfully to a wide variety of problems, including function approximation and classification tasks (Guillén et al., 2006a; Guillén et al., 2005c). Although in the present case the output is discrete, with only two types of pigs, an approximator can be used to obtain a continuous output because Iberian pigs can be crossed with other breeds, and a threshold could be established from a continuous output to offer a more specific classification of the Iberian breed.

One methodology to design an RBFNN consists of a sequence of steps, starting with initialization of the RBF centers. This initialization task has traditionally been solved using clustering algorithms (Guillén et al., 2005a,b; Uykan et al., 2000). The algorithm used in the design of the RBFNNs to classify the different breeds of pig is an adaptation of the algorithm published by Guillén et al. (2006b), in which a fuzzy partition was replaced with a possibilistic-fuzzy partition. This algorithm was used in order to offer a supervised learning in which the output values are considered at the time when centers are placed in one or other positions, allowing identification of a specific class to be prioritized and gaining greater sensitivity and specificity.

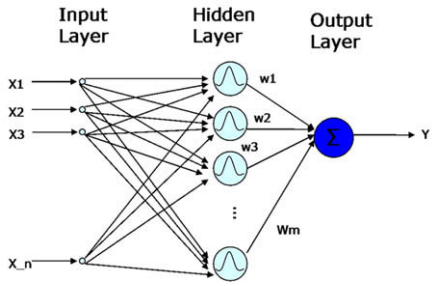

Fig. 1. A radial basis function neural network.

In order to place the centers of RFBs in the input space, a distortion function is defined, which must be minimized to establish a balance between the positions of the centers and the membership degree of the output vectors. This function is defined as:

$$J_{\mathrm{hf,hp}}(U, C, T, O; X) = \sum_{k=1}^{n} \sum_{i=1}^{m} (u_{ik}^{(\mathrm{hf})} + t_{ik}^{(\mathrm{hp})}) D_{ik\mathrm{W}}^2,$$

restricted to the constraints:

$$\sum_{i=1}^{m} u_{ik} = 1, \quad k = 1, \ldots, n$$

$$\sum_{k=1}^{n} t_{ik} = 1, \quad i = 1, \ldots, m,$$

where u_{ik} is the fuzzy membership degree, t_{ik} is the possibilistic membership degree, $D_{ik\mathrm{W}}$ is the weighted Euclidean distance from the center $\vec{c}_i$ to the vector $\vec{x}_k$, and hf and hp are two values to control how fuzzy and possibilistic the partition is.

The Euclidean distance is weighted by defining a hypothetic value for the output axis that assigns an output value o_i to each center. Thus, the weighted Euclidean distance is computed as follows:

$$D_{ik\mathrm{W}} = \|\vec{x}_k - \vec{c}_i\| \cdot |Y_k - O_i|.$$

The final position of the centers is reached by an alternating optimization approach where all elements defined in the function to be minimized are updated iteratively using the equations obtained by differentiating $J_{\mathrm{h}}(U, T, C, O; X)$ with respect to u_{ik}, t_{ik}, $\vec{c}_i$ and, o_i these are:

$$u_{ik} = \left(\sum_{j=1}^{m} \left(\frac{D_{ik}}{D_{jk}} \right)^{\frac{2}{\mathrm{hf}-1}} \right)^{-1}, \quad t_{ik} = \left(\sum_{j=1}^{n} \left(\frac{D_{ik}}{D_{ij}} \right)^{\frac{2}{\mathrm{hp}-1}} \right)^{-1}$$

$$\vec{c}_i = \frac{\sum_{k=1}^{n} (u_{ik}^{\mathrm{hf}} + t_{ik}^{\mathrm{hp}}) \vec{x}_k}{\sum_{k=1}^{n} (u_{ik}^{\mathrm{hf}} + t_{ik}^{\mathrm{hp}})}, \quad O_i = \frac{\sum_{k=1}^{n} (u_{ik}^{\mathrm{hf}} + t_{ik}^{\mathrm{hp}}) y_k d_{ik}^2}{\sum_{k=1}^{n} (u_{ik}^{\mathrm{hf}} + t_{ik}^{\mathrm{hp}}) d_{ik}^2}$$

.

The algorithm continues iterating until the value of the centers does not significantly change. Once the centers have been initialized, the heuristic of the nearest neighboring Ks is used to calculate the radii [knn], and a local search algorithm is applied to make a better fit of the centers and radii.

3. Results and discussion

3.1. Results of macroscopic observations

Macroscopic observations of the M. *masseter* muscle revealed a high number of structured streaks of fat alongside IMCT on the marbled side. This complicates the spectral reflectance measurement, which is influenced by the area on which the VIS/NIRS probe is placed. The streaks can be almost 1 cm thick; hence the reflectance spectrum acquired can be affected by sampling error.

In contrast, macroscopic observation of the lean side of M. *masseter* showed high histological homogeneity, with absence of large streaks of IMCT and fat, revealing a very homogeneous tissue due to the predominance of muscle fibers.

As detailed below, spectral reflectance results obtained on the lean side were more relevant to the classification between Duroc and Iberian pork compared with results obtained on the marbled side.

3.2. Mutual information ranking

The ranking (see Fig. 2) derived from the mutual information ranking (MI Ranking) included 4302 variables for each muscle (2151 wavelengths/spectrum on each side), from which the 1000 wavelengths of greatest importance were selected. Out of these 1000 variables, 951 corresponded to the lean side of M. *masseter* and the remaining 49 wavelengths to the marbled side, with considerably more MI on the lean than on marbled side. Therefore, a much higher amount of information can be obtained from the lean side than from the marbled side, making the lean side much more relevant for training the models that will classify input samples. In Fig. 2, a higher value for a variable indicates a higher relevance for classifying the output. These 1000 variables were correlated with the ten peaks observed after transforming reflectance spectra into absorbance (Absorbance = Log 1/Reflectance), obtaining 10 wavelengths related to these peaks and presenting the highest MI ranking (Table 1).

Fernández-Cabanás et al. (2007) reported that NIR spectra of pork fats are characterized by sharp peaks susceptible to changes caused by unexpected variations, concluding that optimization of the derivatives used is therefore essential to implement robust calibrations for quantitative classifications. In our model, we did not use VIS/NIRS for quantification of the muscle characteristics but rather for qualitative classification of the samples, in line with the model proposed by Cen and He (2007) for pattern recognition by artificial neural networks (ANN). To avoid interference with the analysis capacity of the algorithm and any loss of spectrum information, the MI Ranking was obtained from raw spectrum data with no treatment or mathematical transformation.

3.3. VIS/NIR spectral characterization (VIS/NIR SC)

Figs. 3 and 4 depict mean values of all raw reflectance spectra for the Duroc and Iberian pork and the corresponding absorbance values derived from our study. Reflectance measurements show a clear dominance of the pigments of fresh meat, since muscle fibers, myofibrils, intramuscular fat patches, collagen, protein coagulation surfaces, and moisture cause light scattering (Cozzolino et al., 2003). Only whole muscles were examined in the present study, i.e., with no mincing of samples. As reported by Cozzolino et al. (2003), the structure of muscle appears to influence the way in which light is trapped or scattered and probably interferes with the results. However, this was not sufficient reason for the mincing of samples for their quantification, which is highly incompatible with rapid on-line working, except when the preparation is in

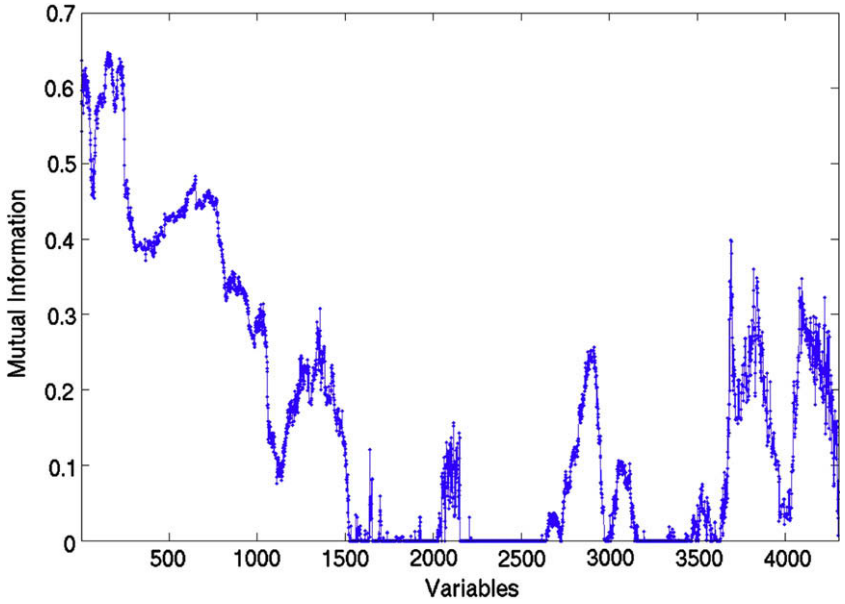

Fig. 2. Mutual information of the 4302 study variables. Variables 1–2151: lean side; variables 2152–4302: marbled side.

Table 1
The most important wavelengths obtained by Mutual Information or VIS/NIRS spectra characterization for the classification of Duroc and Iberian pork, with their respective mutual information rankings and amount of mutual information and the corresponding side of M. *masseter*

Mutual information ranking (set A)				VIS/NIRS spectra characterization (set B)			
Wavelength (nm)	MI ranking	Side	Amount of mutual information	Wavelength (nm)	MI ranking	Side	Amount of mutual information
500	1	Lean	0.6471	414	260	Lean	0.4672
353	14	Lean	0.6368	545	180	Lean	0.5771
996	232	Lean	0.4833	577	53	Lean	0.6246
1073	280	Lean	0.4647	763	646	Lean	0.4023
955	288	Lean	0.4608	983	248	Lean	0.4706
1116	351	Lean	0.4544	1207	821	Lean	0.3532
1888	660	Marbled	0.3981	1446	2003	Lean	0.1372
2018	814	Marbled	0.3597	1722	1279	Lean	0.2459
2292	832	Marbled	0.3469	1928	2858	Marbled	0.0142
2422	957	Marbled	0.3220	2307	3729	Marbled	0
			Σ = 4.5738				Σ = 3.2011

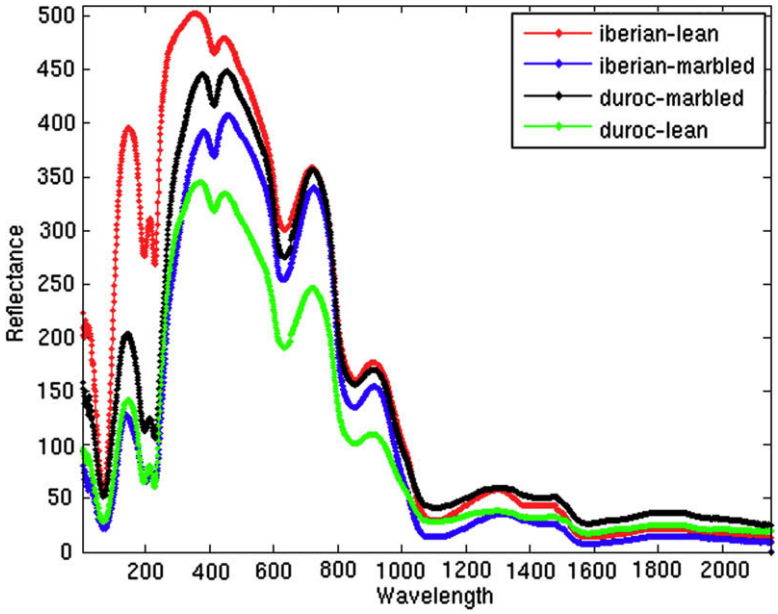

Fig. 3. Mean of reflectance values in the VIS/NIR range of the spectrum. 1: Duroc pork, lean side. 2: Iberian pork, lean side. 3: Duroc pork, marbled side 4: Iberian pork, marbled side.

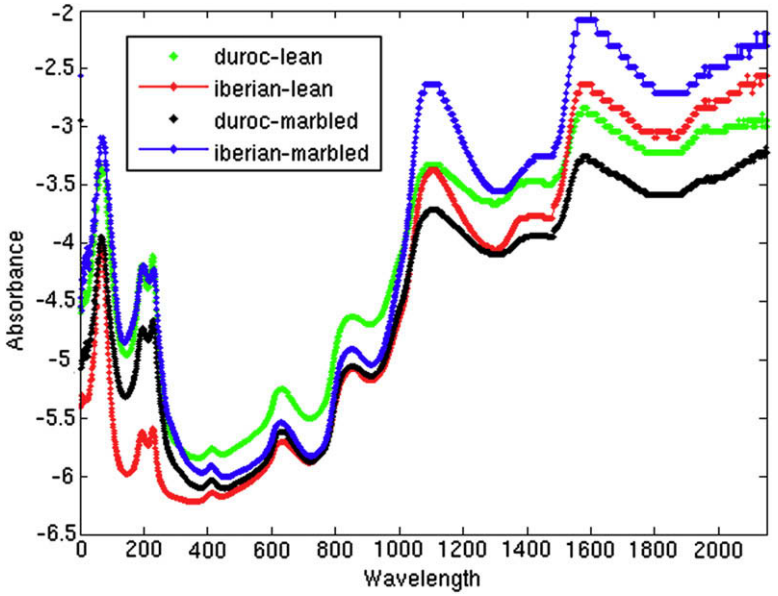

Fig. 4. Mean of absorbance values in the VIS/NIR range of the spectrum. 1: Duroc pork, lean side. 2: Iberian pork, lean side. 3: Duroc pork, marbled side 4: Iberian pork, marbled side.

the form of pork sausages or similar products (Ortiz-Somovilla et al., 2005; Ortiz-Somovilla et al., 2007).

Whereas raw reflectance spectrum data were used for the MI computer analysis, absorbance data were used to characterize the spectra, because they are easier to visualize and are more widely reported in the literature. Fig. 4 depicts the mean VIS/NIR spectra of absorbance in M. *masseter* of Duroc and Iberian pork. The absorbance peaks obtained in our investigation correspond to the main peaks reported in the literature.

In our study, we were able to qualitatively observe a peak of 415 nm in the visible region, associated in the reviewed literature with the Soret Band, which is due to erythrocyte hemoglobin traces (Brøndum et al., 2000; Cozzolino et al., 2003; Cozzolino and Murray, 2004), and adsorption bands at 545 and 577 nm related to muscle pigments, specifically to different hemoglobin oxidation states (Brøndum et al., 2000; Cozzolino et al., 2003; Cozzolino and Murray, 2004; Ortiz-Somovilla et al., 2005; Ortiz-Somovilla et al., 2007). According to Shackelford et al. (2004), the ratio of reflectance at 610 and 525 nm in beef was an indicator of the percentage of myoglobin in an oxymyoglobin state. Their variance component analysis showed that less than half of the total variance was attributable to the carcass at 606 nm because of the large effect of bloom time on reflectance at this frequency. The authors concluded bloom time would have to be standardized if VIS/NIRS spectroscopy were to be successfully used to predict meat quality. The slaughtering process was fully standardized in the

slaughterhouse used in the present study, and the pigs were exsanguinated immediately after their slaughter.

Shackelford et al. (2004) detected a reflectance peak at 804 nm in lamb M. *longissimus dorsi et lumborum* samples using a high-intensity reflectance probe. We used a similar probe from the same company (Analytical Spectral Devices, Inc.) obtaining a reflectance peak of 795–804 nm in both Duroc and Iberian pork. Although the above authors reported that reflectance was not affected by bloom at 804 nm, we were unable to relate this wavelength to any characteristic of the meat.

At 980 nm, we found a new absorbance peak, attributed by Ortiz-Somovilla et al. (2007) to the transition of the spectrum from the visible to near-infrared zone. This peak is due to OH bond second overtone, which is in turn correlated with absorbance at 1450 nm (OH bond second overtone), both due to water (Brøndum et al., 2000; Cozzolino et al., 2003; Hoving-Bolink et al., 2005; Ortiz-Somovilla et al, 2005; Ortiz-Somovilla et al., 2007). An absorption band at 1725 nm was related to CH_2 first overtone (Cozzolino and Murray, 2004) and 1930 nm to OH overtone combinations (water). For fat content and fatty acid molecule, in which a CH bond is essential, researchers have obtained bands at around 1200, 1400, 1750, 2310 and 2340 nm. Various authors (Cozzolino and Murray, 2004; Hoving-Bolink et al., 2005; Ortiz-Somovilla et al, 2005; Ortiz-Somovilla et al., 2007) reported maximum absorbance at 1215 and 1210 nm (CH stretch second overtone), confirming that significant information on fat and intramuscular fat is found at around 1200 nm, as at 1210 nm in the present study. At 2310 nm, CH combinations associated with fat content and fatty acids are observed (Cozzolino et al., 2003; Cozzolino and Murray, 2004).

Above 1800 nm, the crossing of VIS/NIR SC information with MI Ranking data showed that the marbled side was more important, although with a lower amount of MI. The presence of fat on the marbled side of M. *masseter* is evident and more detailed study of these wavelengths is required.

3.4. Breed classification by neural network: Mutual information ranking versus VIS/NIRS spectra characterization

Table 2 shows the comparison between the most relevant wavelengths according to the MI Ranking (set A) and the VIS/NIR SC (set B) obtained from our spectra and the literature. The comparison for breed classification was performed by neural network, carrying out tests using different sizes for the sets of training/test cases of the algorithm (29/1, 28/2, 27/3, 25/5, 26/4 and 20/10). A cross-validation was performed for each of these values. For all these training/test sets, alternative iterations were performed between Duroc and Iberian pigs. Thus, for set 27/3, 2 Duroc pigs and 1 Iberian pig were iterated in the first iteration, 2 Iberian pigs and 1 Duroc pig in the second iteration, and so on. In order to be as equitable as possible and to maintain a balance between classes of samples in the sets, the data were previously randomly assigned such that sample $\vec{x}_i$ was a different class from samples $\vec{x}_{i-1}$ y $\vec{x}_{i+1}$.

Table 2 lists the results of the classifications made by the different RBFNNs developed using the above-described algorithm. The table contains the mean percentage correct classification in each cross-validation for a test set size and gives the total number of samples of Duroc and Iberian samples that were incorrectly classified. The most important variables in set A (500 nm wavelength) and set B (577 nm) are attributable to the main pigment in meat, i.e., myoglobin. There was a lower percentage of correct classifications with the MI model (set A) than with the VIS/NIR SC model (Set B), despite the larger amount of MI for the former (0.6471) versus the latter (0.6246). The VIS/NIR SC had less MI but showed a better fit with the algorithm data, and the values obtained from the literature were more relevant to the neural network classifica-

Table 2
Neural network classification of Duroc and Iberian pork using mutual information (set A) or VIS/NIRS spectra characterization (set B), entering different numbers of input variables as a function of the most important wavelengths

Test set size	Neural network: 1 input variable						Neural network: 3 input variables						Neural network: 10 input variables					
	Mutual information ranking			VIS/NIRS spectra characterization			Mutual information ranking			VIS/NIRS spectra characterization			Mutual information ranking			VIS/NIRS spectra characterization		
	Set A = {500} (nm)			Set B = {577} (nm)			Set A = {500, 353, 996} (nm)			Set B = {577, 545, 414} (nm)			Set A = {500, 353, 996, 1073, 955, 1116, 1888, 2018, 2292, 2422} (nm)			Set B = {577, 545,414, 763, 983, 1207, 1446, 1722, 1928, 2307} (nm)		
	% Correct	Duroc incorrect	Iberian incorrect	% Correct	Duroc incorrect	Iberian incorrect	% Correct	Duroc incorrect	Iberian incorrect	% Correct	Duroc incorrect	Iberian incorrect	% Correct	Duroc incorrect	Iberian incorrect	% Correct	Duroc incorrect	Iberian incorrect
1	70.00	1	8	90.00	1	2	100.00	0	0	96.67	1	0	90.00	1	2	86.67	1	3
2	83.33	1	4	93.33	1	1	100.00	0	0	96.67	1	0	93.30	0	2	90.00	1	2
3	90.00	1	2	93.33	1	1	100.00	0	0	93.33	1	1	93.30	1	1	93.33	1	1
5	90.00	1	2	93.33	1	1	96.60	1	0	96.67	1	0	86.67	1	3	90.00	0	3
6	83.33	1	4	93.33	1	1	100.00	0	0	96.67	1	0	76.67	5	2	76.67	2	5
10	76.60	1	6	93.33	1	1	96.60	1	0	93.33	1	1	76.67	3	4	73.33	8	0

tion than were those derived from MI Ranking. Misclassified samples of Duroc pork had a value of 1 in all of the training/test assumptions, whereas more Iberian pork samples were misclassified. This difference is explained by the very high purity of the Duroc pigs compared with the Iberian pigs, in which there is a genetic mixture of characteristics of the Iberian breed with those of the Duroc breed. Comparison between the MI and VIS/NIR SC models shows that, using the same neural network classification, the percentage of correct classifications is higher for the VIS/NIR SC model, which obtained a very accurate classification for Iberian pork. Because only 30 pigs were studied, the percentages of correct classification are highly influenced by the ratio between the number of training and test cases, although much better results were always obtained with VIS/NIR SC than with MI Ranking. The VIS/NIR SC yielded an accuracy of more than 90% for a single wavelength value (wavelength 577 nm) in the classification between Duroc and Iberian breeds. Nevertheless, it is difficult to discriminate between the two breeds on the basis of 577 nm wavelength because of its continuous nature, since it is attributable to the myoglobin and many factors can influence its oxidation state, including the type of packing (Mancini and Hunt, 2005). At any rate, the consumer can make a relatively accurate classification with the naked eye, with the color of the meat being the main characteristic.

When three input variables were used for the neural network classification (see Table 2), a very high percentage of correct classifications was obtained with both models. The accuracy of the MI model (set A) was greater than 95% in all training/test assumptions, classifying the Iberian pork with no errors for the three most important wavelengths in the MI ranking, i.e., 500, 353 and 996 nm. With the VIS/NIR SC model, none of these three wavelengths corresponded to the maximum peak absorbance wavelength, although they were included in the three main absorbance peaks according to the VIS/NIR SC model: myoglobin pigment, Soret band, and transition area between the visible and near-infrared areas of the spectrum. In set B, as reported in the previous section, took the maximum wavelengths close to the same points selected by MI but for the VIS/NIR SC model, i.e., 577, 545, and 414 nm. With these three wavelengths, the classification produced by the VIS/NIR SC was worse but much more real than that produced by the MI model. We were working with maximum wavelength peaks derived from the literature and we know that these absorbances are due to specific components of meat. In all training/test assumptions (with the exception of 27/3 and 20/10, when a classification accuracy of 93.33% was obtained), an accuracy of more than 95% was achieved and the classification of the Iberian samples was perfect. It is evident that the small number of animals studied (15 Iberian pigs and 15 Duroc pigs) was a major limitation in the training and test sub-sets, and that a study with a larger number of animals is warranted once the utility of the model is established.

In the neural network classification with 10 input variables, the classification accuracy percentages were much lower and very similar between the MI and VIS/NIR SC models. More Iberian pork samples were misclassified compared with the neural network classification with 1 and 3 variables. This is because neural networks do not function well with so many wavelengths; some of which had a very low or zero amount of MI (see Table 1). This suggests that certain wavelengths contribute no relevant information to the neural network, and it is therefore better to remove them from the system. The sum of the amounts of MI derived from the MI Ranking was 29.02% higher than that from the VIS/NIR SC. Hence, the neural network functions better with the MI Ranking, because the ranking of the selected variables are higher.

Overall, there were more misclassified Iberian than Duroc pork samples (Table 2). Although this appears to contradict the description afforded by the algorithm, in which the centers of the RBFs give priority to the Iberian pork samples for their more precise classification, it really reflects the greater genetic mix of the Iberian breed, which has traditionally been crossed with Duroc or other Duroc breeds to improve their productive characteristics and profitability. This explains the worse classification outcomes for the Iberian pork, since one group comprised pure Duroc pigs and the other group included Iberian pigs with different percentages of Duroc $\times$ Iberian mixture. Further study is required with a larger sample of pigs and a more precise knowledge of the mixture between Duroc and Iberian breeds in order to assess the accuracy of this technology.

Because of our small sample size, the percentage of accurate classifications decreased with an increase in the size of the test set. More errors were also observed when more wavelengths were used to identify each pig; an expected finding since the dimensionality of the problem is considerably increased, exponentially expanding the solution space. The more the solution space grows, the more samples are required to identify the parameters of the network so that it can make an accurate classification.

4. Conclusions

This paper presents a combination of VIS/NIRS techniques and artificial neural network as a method to classify different breeds of pork for application in the food industry. Optical properties of the samples yielded an excellent differentiation between Duroc and Iberian breeds. The results reported here indicate that VIS/NIRS can be useful for the objective identification of pig breed. Artificial neural network based on data derived from VIS/NIRS models, as a function of mutual information or VIS/NIR spectra characterization, correctly classified 95% of samples when three wavelengths were entered as input variables. This approach appears to be an objective and rapid method for the authentication and identification of Duroc and Iberian pork.

Acknowledgements

The authors are especially grateful to María Dolores Rodríguez, Francisca Sáez and Jorge A. Payá, technicians at the Department of Pathology of University of Granada for their contribution. This study was funded by the Corporación Tecnológica de Andalucía, Cooperativa del Valle de los Pedroches (COVAP), Bodegas Campos Catering, CICYT (project TIN2004-01419 and TIN2007-60587), Ministerio de Educación y Ciencia, Infraestructura Científico Tecnológica (MEC-JA-FEDER UNGR 05-23-036) and Junta de Andalucia (Project P07-TIC-02768).

References

Astakhov, S., Grassberger, P., Kraskov, A., Stögbauer, H. 2007. Mutual information least-dependent component analysis: http://www.klab.caltech.edu/ kraskov/ MILCA/ (last access: June 1st, 2007).

BOE, 15 de octubre del. 2001. Real Decreto 1083/2001, de 5 de Octubre, Norma de calidad para el jamón ibérico, paleta ibérica y caña de lomo ibérico elaborados en España.

Bonnlander, B.V., & Weigend, A.S. (2004). Selecting input variables using mutual information and nonparametric density estimation. In Proceedings of the ISANN, Taiwan, 2004.

Brøndum, J., Munck, L., Henckel, P., Karlsson, A., Tornberg, E., Engelsen, S.B., 2000. Prediction of water-holding capacity and composition of porcine meat by comparative spectroscopy. Meat Science 55, 177–185.

Cen, H., He, Y., 2007. Theory and application of near infrared reflectance spectroscopy in determination of food quality. Trends in Food Science & Technology 18, 72–83.

Cozzolino, D., Murray, I., 2004. Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. Lebensmittel-Wissenschaft Und-Technologie-Food Science and Technology 37, 447–452.

Cozzolino, D., Barlocco, N., Vadell, A., Ballesteros, F., Gallieta, G., 2003. The use of visible and near-infrared reflectance spectroscopy to predict colour on both intact and homogenised pork muscle. LWT–Food Science and Technology 36, 195–202.

Fernández-Cabanás, V.M., Garrido-Varo, A., García-Olmo, J., De Pedro, E., Dardenne, P., 2007. Optimisation of the spectral pre-treatments used for Iberian pig fat NIR calibrations. Chemometrics and Intelligent Laboratory Systems 87, 104–112.

González-Martín, I., González-Pérez, C., Hernández-Méndez, J., Álvarez-García, N., 2003. Determination of fatty acids in the subcutaneous fat of Iberian breed swine by near infrared spectroscopy (NIRS) with a fibre-optic probe. Meat Science 65, 713–719.

González-Martín, I., González-Pérez, C., Álvarez-García, N., González-Cabrera, J.M., 2005. On-line determination of fatty acids composition in intramuscular fat of Iberian pork loin by NIRS with a remote reflectance fibre optic probe. Meat Science 69, 243–248.

Guillén, A., Rojas, I., González, J., Pomares, H., Herrera, L.J., Valenzuela, O., Prieto, A., 2005a. A possibilistic approach to RBFN centers initialization. Lecture Notes in Computer Science 3642, 174–183.

Guillén, A., Rojas, I., González, J., Pomares, H., Herrera, L.J., Valenzuela, O., Prieto, A., 2005b. Improving clustering technique for functional approximation problem using fuzzy logic: ICFA algorithm. Lecture Notes in Computer Science 3512, 272–280.

Guillén, A., Rojas, I., Ros, E., Herrera, L.J., 2005c. Using fuzzy clustering technique for function approximation to approximate ECG signals. Lecture Notes in Computer Science 3562, 538–547.

Guillén, A., Rojas, I., González, J., Pomares, H., Herrera, L.J., Fernández, F., 2006a. Multiobjective RBFNNs designer for function approximation: An application for mineral reduction. Lecture Notes in Computer Science 4221, 511–520.

Guillén, A., Rojas, I., González, J., Pomares, H., Herrera, L.J., Prieto, A., 2006b. A Fuzzy-possibilistic fuzzy ruled clustering algorithm for RBFNNs design. Lecture Notes in Computer Science 4259, 647–656.

Herrera, L.J., Pomares, H., Rojas, I., Verleysen, M., Guillen, A., 2006. Effective input variable selection for function approximation. Lecture Notes in Computer Science 4131, 41–50.

Hoving-Bolink, A.H., Vedder, H.W., Merks, J.W.M., de Klein, W.J.H., Reimert, H.G.M., Frankhuizen, R., van den Broek, W.H.A.M., Lambooij, enE., 2005. Perspective of NIRS measurements early post mortem for prediction of pork quality. Meat Science 69, 417–423.

Josell, A., Martinsson, L., Borggaard, C., Andersen, J.R., Tornberg, E., 2000. Determination of RN^{-} phenotype in pigs at slaughter-line using visual and near-infrared spectroscopy. Meat Science 55, 273–278.

Kraskov, A., Stögbauer, H., Grassberger, P., 2004. Estimating mutual information. Physics Review E 69, 066138.

Mancini, R.A., Hunt, M.C., 2005. Current research in meat color. Meat Science 71, 100–121.

MAPA 2004. http://www.mapa.es. (Available on-line: February 19th, 2007).

Ortiz-Somovilla, V., España-España, F., De Pedro-Sanz, E.J., Gaitán-Jurado, A.J., 2005. Meat mixture detection in Iberian pork sausages. Meat Science 71, 490–497.

Ortiz-Somovilla, V., España-España, F., Gaitán-Jurado, A.J., Pérez-Aparicio, J., De Pedro-Sanz, E.J., 2007. Proximate analysis of homogenized and minced mass of pork sausages by NIRS. Food Chemistry 101, 1031–1040.

Ramírez, M.R., Cava, R., 2007. Effect of Iberian X Duroc genotype on dry-cured loin quality. Meat Science 76, 333–341.

Savenije, B., Geesink, G.H., van der Palen, J.G.P., Hemke, G., 2006. Prediction of pork quality using visible/near-infrared reflectance spectroscopy. Meat Science 73, 181–184.

Shackelford, S.D., Wheeler, T.L., Koohmaraie, M., 2004. Development of optimal protocol for visible and near-infrared reflectance spectroscopic evaluation of meat quality. Meat Science 68, 371–381.

Stögbauer, H., Kraskov, A., Astakhov, S.A., Grassberger, P., 2004. Least dependent component analysis based on mutual information. Physics Review E 70, 066123.

Swatland, H.J., 1994. Physical measurements of meat quality: Optical measurements, pros and cons. Meat Science 36, 251–259.

Uykan, Z., Güzelis, C., Celebei, M.E., Koivo, H.N., 2000. Analysis of input–output clustering for determining centers of RBFN. IEEE Transactions on Neural Networks 11, 851–858.

Xia, J.J., Berg, E.P., Lee, J.W., Yao, G., 2007. Characterizing beef muscles with optical scattering and absorption coefficients in VIS–NIR region. Meat Science 75, 78–83.