2nd FOOD CHEMISTRY Conference
Shaping the Future of Food Quality, Safety, Nutrition and Health
17–19 September 2019 • Seville, Spain

NIRS and artificial neuronal network to differentiate "Jamón de Guijuelo" DO Iberian dry ham

Hernández-Ramos, P.[1]; Martínez-Martín, I.[2]; Hernández-Jiménez, M.[2]; Vivar-Quintana, A.M.[2]; Revilla, I[2]; González-Martín, M.I.[3]

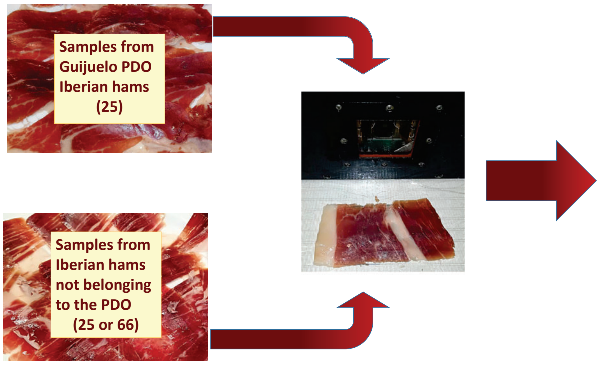
[1] *Graphic Expression in Engineering, University of Salamanca*
[2] *Food Technology, University of Salamanca*
[3] *Analytical Chemistry, Nutrition and Bromatology, University of Salamanca*

Introduction & Aims

In Spain the largest producer of Iberian ham is the Protected Designation of Origin (PDO) of "Guijuelo". The price of these products gives rise to the need for the development of a reliable methodology to allow the authentication of the ham so as to avoid fraud. Near infrared spectroscopy (NIR) is one of the techniques with the potential for this purpose. However, NIR spectral information due to its complexity requires a multivariate data analysis such as artificial neural networks (ANN). In this study spectral information obtained from a NIRS analysis was therefore used to test the authenticity of Iberian ham by means of ANN.

Materials & Methods

In order to do so, 91 ham samples, 25 samples from Guijuelo PDO Iberian hams, and 66 samples from Iberian hams not belonging to the PDO, were analyzed using a NIRS coupled with a fiber optic. Two groups of samples were analyzed: the first one with the same number of samples (25 vs 25), and the second one with a different number of samples (25 vs 66) in the PDO and non PDO groups.



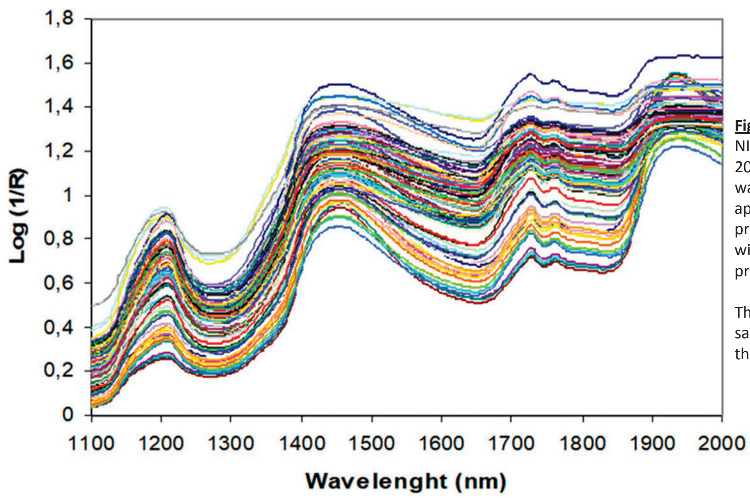


Figure 1. Recording of the NIR spectra from 1100 to 2000 nm at intervals of 2 nm was accomplished by direct application of the fiber optic probe to the ham slice without any previous preparation.

The 451 data from each sample were the inputs of the ANN.

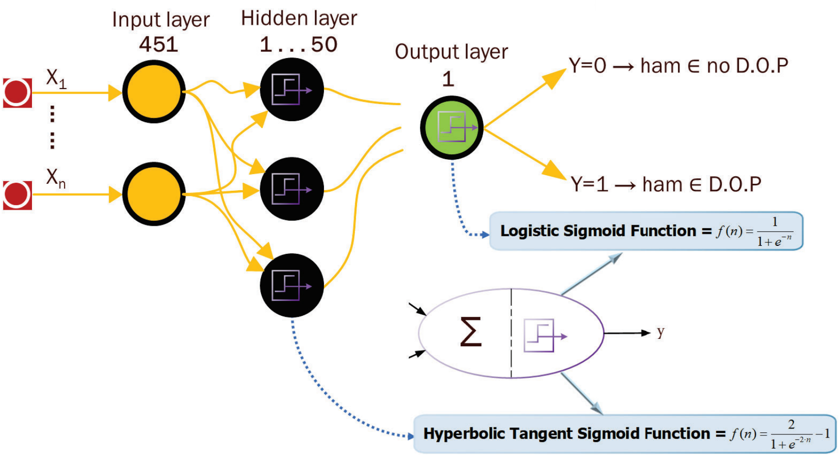


Figure 2. The Neural Network used was the Multi-Layer Perceptron (MLP) feedforward ANN with 451 neurons in the input layer, one neuron in the output layer that shows the assigned category and one hidden layer with a variable configuration of between 1 and 50 neurons. The Hyperbolic Tangent Sigmoid and Logistic Sigmoid transfer functions were used in the hidden and output layers respectively. MatLab R2018 (MathWorks®) software has been used.

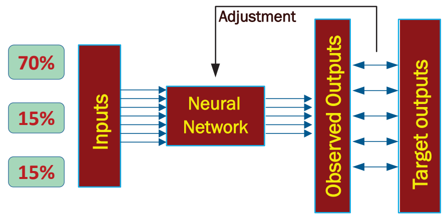


Figure 3. Supervised learning by Backpropagation (BP) error was used for network training. 10 ANN different training algorithms were examined (Table 1).
The original data set (50 or 91 observations depending on the sample group analyzed) was divided at random as follows:

- 70% of the observations were used for network training
- 15% of the observations were used for network validation and for detecting the overfiting
- 15% of the observations were used for testing the accuracy of the trained network

For each number of neurons in the hidden layer, 300 training times were carried out with different inital weights randomly established.
The suitability of the networks obtained was established by minimizing the mean squared errors (MSEs) between the targets and the ANN outputs obtained during training process to establish the connection weights.

Results & Discussion

The results obtained show that some network architectures could be found to allow 100% correct classification during the test process.

The results were obtained after more than 300 trainings. In many cases over 1000 EPOCs were necessary for each of training.

Such a high number of EPOCs may increase training accuracy but this also increases the risk that the model was merely memorizing the training set.

Although the results are promising, when the selected network architectures are applied to the training and validation stages, the success rate decreases owing to the small number of samples used in this test.

Table 1 also shows the tendency of the network to classify the samples in the largest category when the network is trained with an unbalanced number of samples of each category, witch reduces the success rate.

Training algorithm by backpropagation (BP)	Sample Group: 25 vs 25		Sample Group: 25 vs 66	
	% adjustment	Nº hidden layer neurons	% adjustment	Nº hidden layer neurons
Conjugate Gradient BP with Powell-Beale restarts.	100	38	71.43	7
Conjugate Gradient BP with Fletcher-Reeves updates.	100	5	64.29	4
Conjugate Gradient BP with Polak-Ribiere updates.	100	11	71.43	44
Gradient Descent BP.	100	1	100	5
Gradient Descent with adaptive Learning Rate.	100	24	64.29	33
Gradient Descent with momentum.	100	5	64.29	18
Gradient Descent with momentum and adaptive Learning Rate BP.	100	25	64.29	27
One Step Secant BP.	100	16	71.43	14
RPROP Resilient BP.	100	18	71.43	4
Scaled Conjugate Gradient BP.	100	22	57.14	1

Table 1. % of success in the correct classification according to group of samples and number of neurons in the hidden layer for the best network of each training algorithm analyzed.

Conclusions

- The network operates better when the number of samples belonging to each class under study is balanced.
- To make the operation of the networks more stable and efficient, more samples are needed for training process.

ACKNOWLEDGMENTS: We thank the Provincial Council of Salamanca (Spain) for the concession of the project 18VEUH 463AC06 and to Carrasco Ibéricos Company (Guijuelo, Spain)