AUTHOR QUERY FORM

ELSEVIER	**Journal: JFOE** **Article Number: 7697**	**Please e-mail or fax your responses and any corrections to:** **E-mail: corrections.esch@elsevier.sps.co.in** **Fax: +31 2048 52799**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	**Query / Remark: click on the Q link to go** **Please insert your reply or correction at the corresponding line in the proof**
Q1	Please confirm that given name(s) and surname(s) have been identified correctly.
Q2	Reference 'Picouet et al. (2010)' is cited in the text but not provided in the reference list. Please provide it in the reference list or delete this citation from the text.
Q3	The citation 'Holmes et al. (2012)' has been changed to match the date in the reference list. Please check here and in subsequent occurrences, and correct if necessary.
Q4	This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or, alternatively, delete it. Any reference not dealt with will be retained in this section.

Please check this box if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Highlights

• Data mining and MRI–CVT have been firstly used to study quality features of hams. • Data mining tasks are appropriate to deduce and predict quality traits of hams. • Physical–chemical and computer vision data are inferred by applying deductive tasks. • Quality traits can be control by using predictive techniques and computer vision data.

Journal of Food Engineering xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Food Engineering

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

Applying data mining and Computer Vision Techniques to MRI to estimate quality traits in Iberian hams

Trinidad Pérez-Palacios [a,*], Daniel Caballero [b], Andrés Caro [b], Pablo G. Rodríguez [b], Teresa Antequera [a]

[a] Tecnología de los Alimentos, Facultad Veterinaria, Universidad de Extremadura, Av. Universidad s/n, 10003 Cáceres, Spain
[b] Departamento de Ingeniería de Sistemas Informáticos y Telemáticos, Escuela Politécnica, Universidad de Extremadura, Av. Universidad s/n, 10003 Cáceres, Spain

A R T I C L E I N F O

Article history:
Received 9 July 2013
Received in revised form 22 December 2013
Accepted 21 January 2014
Available online xxxx

Keywords:
Data mining
MRI and Computer Vision Techniques
Deduction
Prediction
Quality parameters
Iberian ham

A B S T R A C T

This study aims to forecast quality characteristics of Iberian hams by using non-destructive methods of analysis and data mining. Magnetic Resonance Imaging and Computer Vision Techniques were conducted on hams throughout their processing. Physico-chemical parameters were also measured in these products. Information from these analyses was integrated in a database. First, deductive techniques of data mining were applied to these data. Multiple linear regression allows for the estimation of information from Magnetic Resonance Imaging, Computer Vision Techniques and physico-chemical analysis. This enables the completion of the initial database. Then, predictive techniques of data mining were applied. Both, multiple linear regression and isotonic regression achieved the prediction of weight, moisture and lipid content of hams as a function of features obtained by Magnetic Resonance Imaging and Computer Vision Techniques. Thus, data mining, Magnetic Resonance Imaging and Computer Vision Techniques could be used to estimate the quality traits of Iberian hams. This allows for the improvement of the process control without destroying any piece.

1. Introduction

Quality attributes of dry-cured hams depend on characteristics of raw material and processing conditions. Throughout the processing of hams, changes on the physico-chemical (P-C) characteristics of the thighs take place, also influencing the quality of the final product. Thus, not only characteristics of thighs but also their modifications during the processing are important parameters to control the technological process of dry-cured hams (Pérez-Palacios et al., 2011a).

Temperature and relative humidity conditions during the processing lead to ham dehydration and, hence, to weight loss. The ham industry estimates the optimal ripening time by the percentage of weight loss, related to the amount of water contained in the ham muscles (Martin et al., 1998). Raw material for ham production should contain plenty of intramuscular fat, which is an important characteristic, due to its positive influence on quality parameters on the final product, such us marbling, juiciness, odour, and aroma (Ruiz et al., 2002).

Usual methods for evaluation of the P-C characteristics (i.e. weight loss, moisture, fat content) of dry-cured hams throughout the whole processing are tedious and time-consuming, and sometimes involve the destruction of the pieces. In this sense, the use of non-destructive techniques, such as computed tomography (CT), near infra-red reflectance spectroscopy (NIRs) and Magnetic Resonance Imaging (MRI), has been proposed for determining quality parameters in this product. Studies on salt content by means of CT have been carried out by several authors (Fulladosa et al., 2010; Haseth et al., 2012; Picouet et al., 2013; Santos-Garcés et al., 2010; Vestergaard et al., 2005). CT has also been applied for predicting the water content throughout the process of hams (Fulladosa et al., 2010; Santos-Garcés et al., 2010), and the weight and lean content of the raw material (Picouet et al., 2010). In pig carcass, Furnols et al. (2009) estimated the lean meat content by using CT. Collell et al. (2011) used NIRs to predict moisture, water activity and NaCl content at the surface of dry-cured ham during the process. Results obtained by Pérez-Juan et al. (2010) showed the accuracy of NIRs to predict the fatty acid composition of ham subcutaneous fat.

MRI is a non-destructive, non-invasive, non-intrusive, non-ionizing and innocuous technique. Thus, as an alternative to P-C procedures, MRI has also been proposed to study some characteristics in hams. Fantazinni et al. (2009) used this technique to obtain information on moisture and salt distribution during the

Abbreviations: KDD, Knowledge Discovery in Databases; R, raw hams; SA, end of salting; PS, end of post-salting; D, end of drying; DC, dry-cured hams; P-C, physico-chemical; MRI–CVT, Magnetic Resonance Imaging and Computer Vision Techniques; B, *Biceps femoris muscle*; S, *Semimembranosus muscle*.

* Corresponding author. Tel.: +34 927 257123; fax: +34 927 257110.
E-mail address: triny@unex.es (T. Pérez-Palacios).

0260-8774/$ - see front matter © 2014 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.jfoodeng.2014.01.015

processing of Parma hams. Recently, predictive models have been proposed for estimating water activity, moisture, salt content and proteolysis extent in S. Daniele hams on the basis of the MR signal intensity (Manzoco et al., 2013).

The implementation of active contours in MRI can be used to recognize the *Biceps femoris* and *Semimembranosus* muscles in Iberian hams, determine the volume of the muscle and estimate ham weight and moisture (Antequera et al., 2007; Caro et al., 2001). MRI and computational texture features allowed for the classification of fresh and dry-cured Iberian hams as a function of pig feeding background (Pérez-Palacios et al., 2010a, 2011b). Sensory traits in Iberian dry-cured hams were predicted from computational texture characteristics obtained from MRI of fresh hams (Pérez-Palacios et al., 2010b).

The calculation of intramuscular fat levels of Iberian ham has also been attempted by using MRI applications (Ávila et al., 2005; Caro et al., 2003), obtaining reasonable, but not very high, correlation coefficients (around 0.50–0.63), which shows the potential of this technique for determining intramuscular fat level in Iberian hams. In these studies, database obtained from P-C, and MRI and computational analysis are processed by applying usual statistical tools such as Pearson's correlation coefficients or principal components analysis (Pérez-Palacios et al., 2010b, 2011b). The integration of heterogeneous P-C information with computer vision data, and the analysis of this new data set by data management and database applications would be innovative and could give accurate results, playing an increasing role in furthering food research (Cortez et al., 2009; Holmes et al., 2007).

Data mining is an important part of a larger process known as KDD ("Knowledge Discovery in Databases") (Fayyad et al., 1996). It is associated with large data. The main goal of data mining consists in extracting hidden information from a data set. This can be achieved by the automatic or semi-automatic analysis of large amounts of data, which allows for the extraction of interesting and previously unknown patterns (Hastie et al., 2001). These patterns can be groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies among data (association rules). Thus, the patterns can be seen as a summary of the input data, and can be used for further analysis.

Interest in data mining has recently grown because of the rapidly decreasing cost of large storage devices and increasing ease in data collection over networks. Other factors include, the development of robust and efficient algorithms to process this data, and the increase in computing power, enabling the use of intensive computational methods for data analysis (Mitchell, 1999).

To our knowledge, few studies apply data mining to food. Song et al. (2002) and Cortez et al. (2006) used this computing technique to predict quality traits in beef and lamb, respectively. It has also been used to predict the oxidation of menhaden fish oil (Klaypradith et al., 2010) or to model wine preferences (Cortez et al., 2009). Holmes et al. (2007) applied data mining to detect fruit and vegetables contaminated with pesticide and to identify these products as a function of their home country.

For this study, data obtained from the MRI–CVT (volume) and P-C analysis (moisture, lipid and weight) of a homogeneous Iberian ham batch were used to construct a database. Several data mining techniques were applied to this database in order to (i) estimate values for the analysed parameters in a higher number of samples and (ii) predict moisture, lipid content and weight throughout the processing of the Iberian ham.

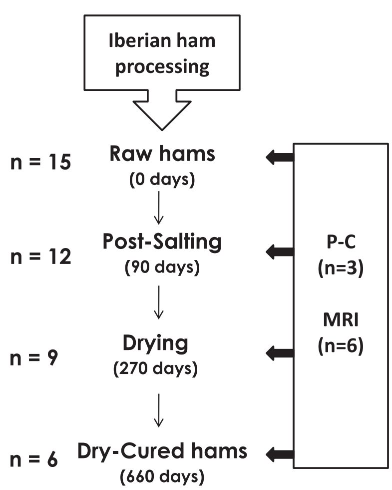

Fig. 1. Sampling throughout the Iberian ham processing for the physico-chemical analysis (P-C) and the MRI acquisition.

2. Material and methods

2.1. Experimental design

This study was carried out with 15 Iberian thighs, which were processed following the traditional processing as described in Antequera et al. (2007). Four stages were considered: raw hams (R), 0 days; end of post-salting (PS), 90 days; end of drying (D),

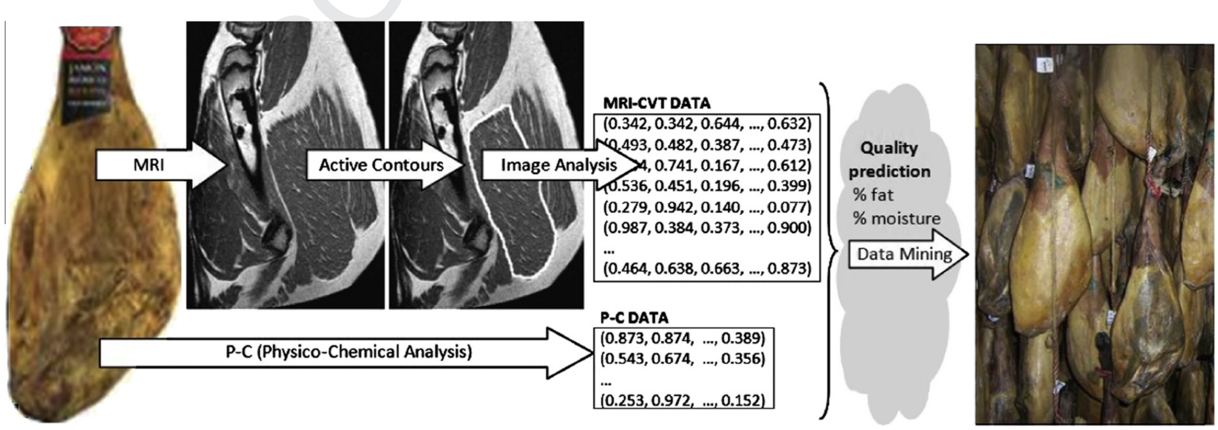

Fig. 2. Acquisition of Iberian ham data (from physico-chemical analysis and MRI and Computer Vision Techniques) used to estimate quality parameters by applying data mining.

A

N	Stage	HW	BW	BM	BL	SW	SM	SL	HV	BV	SV
3	1	10,800									
4	1	11,000									
8	1	11,000									
12	1	11,200							80,542	19,614	24,639
13	1	11,200									
17	1	10,800							78,728	20,838	23,785
19	1	10,600									
22	1	11,200									
23	1	11,000							82,764	20,426	24,571
24	1	10,600	1,235	69.16	9.18	725	72.57	4.14			
27	1	10,800							84,275	19,025	23,073
28	1	11,000							82,220	20,114	31,994
32	1	11,200							80,624	20,856	25,150
34	1	11,000	1,435	64.97	9.52	755	70.36	4.24			
37	1	11,000	1,475	71.95	7.61	760	72.68	3.46			
3	1.5	10,600									
4	1.5	10,800									
8	1.5	10,800									
12	1.5	11,000									
13	1.5	10,900									
17	1.5	10,500									
19	1.5	10,300									
22	1.5	10,900									
23	1.5	10,800									
27	1.5	10,600									
28	1.5	10,800									
32	1.5	11,000									

⋮

B

N	Stage	HW	BW	BM	BL	SW	SM	SL	HV	BV	SV
3	1	10,800	1,391	69.37	8.49	728	69.92	6.18	68,276	20,244	23,632
4	1	11,000	1,388	69.17	8.32	724	73.19	6.27	71,383	20,441	23,176
8	1	11,000	1,386	69.15	8.32	715	72.80	6.27	73,473	20,315	23,704
12	1	11,200	1,392	69.14	7.92	712	73.96	6.40	80,542	19,614	24,639
13	1	11,200	1,393	69.39	8.01	733	72.49	6.16	79,040	20,354	25,425
17	1	10,800	1,378	69.27	8.66	724	70.56	5.95	78,728	20,838	23,785
19	1	10,600	1,323	69.29	8.99	720	69.82	6.12	76,628	20,089	23,866
22	1	11,200	1,402	68.99	8.10	721	73.89	5.99	83,822	20,123	26,180
23	1	11,000	1,392	69.36	8.22	731	71.20	6.14	82,764	20,426	24,571
24	1	10,600	1,235	69.16	9.18	725	72.57	4.14	74,918	20,384	22,895
27	1	10,800	1,385	68.96	8.12	695	70.63	6.29	84,275	19,025	23,073
28	1	11,000	1,380	69.08	9.06	732	71.32	5.64	82,220	20,114	31,994
32	1	11,200	1,382	69.01	8.20	749	72.68	6.19	80,624	20,856	25,150
34	1	11,000	1,435	64.97	9.52	755	70.36	4.24	84,922	19,949	26,507
37	1	11,000	1,475	71.95	7.61	760	72.68	3.46	92,725	20,027	27,599
3	1.5	10,600	1,296	65.68	8.78	664	67.33	6.58	69,061	18,083	21,587
4	1.5	10,800	1,330	64.24	9.44	670	68.05	6.57	72,089	18,680	21,736
8	1.5	10,800	1,315	63.98	9.44	662	67.91	6.54	74,128	18,643	22,272
12	1.5	11,000	1,262	65.87	9.01	702	70.82	6.41	76,602	18,497	23,505
13	1.5	10,900	1,442	66.11	8.96	678	69.55	6.66	72,824	18,214	23,983
17	1.5	10,500	1,246	65.20	8.99	656	65.24	6.48	75,279	18,044	22,087
19	1.5	10,300	1,279	64.95	9.72	654	65.09	6.27	73,231	18,449	22,648
22	1.5	10,900	1,290	65.05	8.98	661	68.61	6.57	83,029	18,491	24,184
23	1.5	10,800	1,347	65.23	9.79	689	67.58	6.91	79,061	18,512	24,112
27	1.5	10,600	1,271	65.45	9.39	682	67.11	6.28	80,382	18,479	24,588
28	1.5	10,800	1,285	65.18	9.05	674	66.86	6.31	85,031	18,628	25,043
32	1.5	11,000	1,292	65.04	8.91	657	68.90	6.56	89,310	18,470	25,793

⋮

Fig. 3. Initial database with incomplete records (A) and with all records filled after applying data mining (B). The data set of each record is composed by (i) processing stages (Stage) (raw hams = 1; salting = 1.5; post-salting = 2; drying = 3; dry-cured ham = 4), (ii) physico-chemical parameters (ham weight = HW, *Biceps femoris* muscle weight, moisture and lipid content = BW, BM and BL, respectively, *Semimembranosus* muscle weight, moisture and lipid content = SW, SM and SL, respectively), and (iii) MRI and Computer Vision Techniques (ham, *Biceps femoris* and *Semimembranosus* volume = HV, BV, and SV, respectively). Suspension dots indicate that the database is greater and it has been cut. N = ham identifier. HW, BW and SW are expressed in grams; BM, BL, SM and Sl are expressed in g/100 g sample; HV, BV and SV are expressed in voxel.

270 days; and dry-cured hams (DC), 660 days. This experimental design is shown in Fig. 1.

At each stage, 6 hams were scanned for obtaining MR images. After then, three hams were destroyed at each stage for the P-C analysis, having 12, 9 and 6 hams at PS, D and DC stages, respectively. Ham weights were recorded at these four stages and also at the end of the salting step (SA). This is how the P-C data set is formed.

In this work MRI has been used as a non-invasive technique only to acquire images of the hams without destroying them. Then, our own active contour algorithms were applied to recognize the *Biceps femoris* and *Semimembranosus* mucles, in order to compute their volumes, as described in Antequera et al. (2007).

Numerical data is extracted by Data Mining from the data sets obtained by our MRI–CVT and from the data sets obtained by P-C. Fig. 2 describes the whole process.

2.2. MRI acquisition

Magnetic resonance images were generated at the "Infanta Cristina" University Hospital (Badajoz, Spain). A MRI scanner (Philips Gyroscan NT Intera 1.5 T) was used, with a quadrature whole-body coil. Sequences of T1 were applied with the following parameters: 120 × 85 mm for field-of view (FOV), 20 ms for echo time (TE), 500 ms for repetition time (TR), 2 mm thick slices, 90° for flip angle, i.e. a T1-weighted spin echo (SE), 0.23 × 0.20 mm per pixel resolution. Sixty slices per ham piece were obtained. The MRI acquisition was done at 20 °C and it took 28 min for each ham. All the images were in DICOM format, with a 512 × 512 resolution, and 256 grey levels.

2.3. Computer Vision Techniques

After the images were acquired, our own computer vision algorithms were applied to extract numerical data from these images. Then, data mining techniques were tested over these data to obtain prediction equations.

The automated procedure was run as described in Fig. 2. First, a previous image pre-processing stage was carried out. Then, the *Biceps Femoris* and *Semimembranosus* muscles (B and S, respectively) were recognized distinctly by using Active Contours, applying a greedy algorithm method (Antequera et al., 2007). The surface and volume for all the contours is calculated by relying on classical methods in analytical geometry. Volume is expressed in voxel (*volume per element*), which is 0.23 × 0.2 × 2 mm^3.

2.4. Physico-chemical analysis

At each stage of the processing, ham weight was recorded and the B and S muscles of three hams were dissected, weighed and analysed for moisture (AOAC, 2000; reference 935.29) and lipid content (Pérez-Palacios et al., 2008a). Analyses were done in triplicate.

2.5. Data mining

The free software WEKA (Waikato Environment for Knowledge Analysis) (http://www.cs.waikato.ac.nz/ml/weka/) was used for carrying out the data mining analysis. The primary groups in data mining tasks are descriptive and predictive techniques. The first ones include deductive techniques, which have the ability to infer new values based on actual data. In predictive techniques, future models can be predicted from current data by trend analysis (Witten and Frank, 2005; Wu et al., 2008). Both, descriptive and predictive techniques were applied in this study.

Multiple linear regression was used for the deductive tasks. The dependent variable to be estimated was always unique and numerical and this method enables the removal of collinear attributes. In addition, regression techniques seem to be the most appropriate to forecast values, as it allows inferring numerical data from the available numerical values. The M5 method of attribute selection and a ridge value of 1×10^{-4} were applied. This method steps through the attributes, and removes the one with the smallest standardised coefficient until no improvement is observed in the estimate of the error given by the Akaike information criterion (Hastie et al., 2001).

Again, multiple linear regression was used for the experiments of prediction. This technique obtains a linear regression equation, which can be used to predict future values (Hastie et al., 2001). The M5 method of attribute selection and a ridge value of 1×10^{-4} were also applied.

Isotonic regression was also tested for prediction. When the values of the database are highly correlated, the use of non-linear regression is recommended. In these cases, the isotonic regression is considered as a good option. Isotonic regression provides a set of values from the information stored on a database. It is based on estimating ordered values for an independent variable (i.e. weight) as a function of one of the input parameters (attributes of the database). Thus, the ham weight is predicted as a function of the volume or the maturation stage. Only the input parameters providing better adjustment results (for example, the stage) will be selected. Finally, an interpolation function is established (polynomial trend line) to compare the provided set data with original values in the database, obtaining the prediction equation (Borge, 1985; Barlow et al., 1972).

2.6. Databases

An initial database was built with data obtained throughout the ham processing: (i) stage of the ham processing, (ii) P-C analysis (ham, B and S weight; moisture and lipid content of the B and S), and (iii) MRI–CVT (ham, B and S volume) (see Fig. 3).

As previously explained (Fig. 1), this study was carried out with 15 Iberian hams and three of them were discarded at each stage. Thus, the number of pieces at R, SA, PS, D and DC stages were 15, 12, 12, 9 and 6, respectively. The initial database contained 54 records, with each record treated as a data set obtained from a ham. Although this database might be regarded as small, it should be noted that each Iberian ham presents considerable costs, about 30 Euros per kilo plus lab work.

Since the 15 hams were not analysed at all the ripening stages, this initial database presents incomplete records (Fig. 3A). After applying data mining techniques (multiple linear regression), the values for all analysed parameters were estimated. The records thus completed made up the whole database, as can be observed in Fig. 3B.

2.7. Statistical design

Differences throughout the processing of Iberian hams with parameters determined by P-C analysis and MRI–CVT were analysed by one-way analysis of variance (ANOVA). When significant differences ($p < 0.05$) were found, the Tukey's test was conducted. Analyses were done by using the SPSS package (v.18.0).

3. Results and discussion

3.1. Physico-chemical and MRI-Computer Vision Techniques

Table 1 shows results on ham weight, moisture content, lipid content, and weight of B and S muscles in Iberian hams throughout the processing. Weight and lipid content in B are known to be

Table 1
Results on physico-chemical analysis (ham weight = HW, *Biceps femoris* muscle weight, moisture and lipid content = BW, BM and BL, respectively, *Semimembranosus* muscle weight, moisture and lipid content = SW, SM and SL, respectively) at the different stages of the Iberian ham processing.[a]

	HW (g)	BW (g)	BM (g/100 g sample)	BL (g/100 g sample)	SW (g)	SM (g/100 g sample)	SL (g/100 g sample)
Raw hams	10960 ± 203a	1382 ± 129a	68.69 ± 3.17a	8.77 ± 0.98b	747 ± 19a	71.87 ± 1.34a	3.95 ± 0.45
Salting	10750 ± 211a	NA	NA	NA	NA	NA	NA
Postsalting	9683 ± 301b	1130 ± 30b	60.66 ± 1.16b	11.57 ± 1.79b	527 ± 15b	60.37 ± 1.65b	5.53 ± 0.32
Drying	8489 ± 401c	1030 ± 83b	54.43 ± 0.66c	10.63 ± 1.11b	552 ± 53b	34.24 ± 5.04c	6.66 ± 1.05
Dry-cured hams	7700 ± 110d	713 ± 21c	42.92 ± 2.49d	16.94 ± 1.44a	327 ± 70c	25.71 ± 2.76d	6.27 ± 1.99
p	<0.001	<0.001	<0.001	0.002	<0.001	<0.001	0.2

NA = not analysed.
In the same column, means with different letters differ significantly between stages.
[a] Values are expressed as means ± standard deviation.

greater in comparison to the S muscle (Pérez-Palacios et al., 2008b, 2010c), which is corroborated in this study. As expected, ham and muscle weight and moisture decreased during the processing due to water loss (Martin et al., 1998; Pérez-Palacios et al., 2011b). The significant increase in the percentage of lipid content in the B muscle during the processing can be also related to the water loss, since the percentage of dry matter (as fat) increased as the water content decreased.

Moisture loss during the processing occurs more in the S muscle than in B, above all at the last stages of the processing. This fact agrees with previous results in Andres et al. (2005). This phenomenon is related to muscle location in the ham (the B muscle is an internal muscle, while S is external), since water loss takes place from the inner to the outer part. Thus, water loss is facilitated in external muscles, such as S.

Volume of ham, B and S muscles at R, PS, D and DC stages achieved by MRI–CVT are shown in Table 2. These three objects of study decreased during processing, which coincides with the changes found in ham and muscle weight. The accuracy of volume estimation for the muscles is very high, as can be examined in Antequera et al. (2007). There was a high correlation ($R^2 = 0.992$) between the data obtained by physical measurement and sizes measured on MRI by computer vision methods.

3.2. Data mining for deduction

As previously explained, a database with 54 records was built (Fig. 3A). A record is the data set of a ham, which includes (i) the stage of the processing, (ii) data from P-C analysis (ham weight, B and S muscles weight, moisture and lipid content), and (iii) data from MRI–CVT (ham, B and S muscles volume). Most of the records in the database were incomplete. By applying multiple linear regression, the unknown information of the records in the database is estimated. Hence, a database of 54 full records is computed (Fig. 3B). This process could be seen as a type of data reconstruction: data that did not exist is reconstructed by using various algorithms with some degree of confidence.

Correlation index R^2 is used to prove the correctness and precision of the estimated values by using multiple linear regression. Table 3 shows the correlation coefficients between real and predicted data for the features analysed: ham weight; B and S muscles weight, moisture and lipid content; ham, B and S muscle volume. As can be seen, high correlations ($R^2 > 0.900$) have been obtained for all traits, except for lipid content of the S muscle ($R^2 = 0.665$). This lower correlation could be related to the high variability of fat content in Iberian ham. Particularly noteworthy is the high correlation obtained for moisture in the two muscles (>0.990).

Table 4 displays the value range of the predicted features, which can be compared to the real values shown in Table 1, for the P-C characteristics, and Table 2, for data obtained by MRI–CVT, in order to corroborate the good correlation between real and predicted data. For example, at the R stage, the average moisture of the B muscle was 68.69% (BM value at Raw in Table 1) and the values predicted for this characteristic range between 64.97% and 71.95% (BM value at Raw in Table 4); at the D stage, the real value for ham volume was 64.50 voxel, and its predicted values were 58.94–67.21 voxel.

To the best of our knowledge, deductive methods from data mining techniques have not been applied at all in food science. This fact is really important since this approach yields a large number of data from a small and incomplete database. In the case of Iberian ham production, the application of deductive methods of data mining would be an interesting tool due to the high cost of this product.

3.3. Data mining for prediction

The prediction of ham quality parameters (weight, moisture content, and lipid content in the B and S muscles) was also tested. Predictive techniques from data mining were applied to informa-

Table 2
Results on MRI and Computer Vision Tecniques (ham, *Biceps femoris* and *Semimembranosus* volume = HV, BV, and SV, respectively) at the different stages of the Iberian ham processing.[a]

	HV (voxel)	BV (voxel)	SV (voxel)
Raw hams	81520 ± 1950a	25530 ± 7200a	25530 ± 3240a
Salting	NA	NA	NA
Postsalting	75250 ± 2050b	21640 ± 8900b	21640 ± 1070b
Drying	64500 ± 2170c	15146 ± 1230c	15140 ± 1730c
Dry-cured hams	56990 ± 5630d	12130 ± 1270d	12130 ± 1710c
p	<0.001	<0.001	<0.001

NA = not analysed.
[*] Values are expressed as means ± standard deviation.

Table 3
Correlation coefficient (R^2) between real and predicted data obtained by data mining for the features analysed by physico-chemical analysis and MRI and Computer Vision Techniques.

	R^2
BW	0.975
SW	0.916
BM	0.994
SM	0.993
BL	0.908
SL	0.665
HV	0.975
BV	0.999
SV	0.993

See abbreviations in Fig. 3.

Please cite this article in press as: Pérez-Palacios, T., et al. Applying data mining and Computer Vision Techniques to MRI to estimate quality traits in Iberian hams. Journal of Food Engineering (2014), http://dx.doi.org/10.1016/j.jfoodeng.2014.01.015

Table 4
Minimum and maximum values for the features predicted by using data mining.

	BW (g)	BM (g/100 g)	BL (g/100 g)	SW (g)	SM (g/100 g)	SL (g/100 g)	HV (voxel)	BV (voxel)	SV (voxel)
Raw hams	1235–1475	64.97–71.95	7.61–9.52	695–760	69.82–73.96	3.46–6.40	68270–92720	19020–20850	22980–31990
Salting	1245–1442	63.98–66.11	8.78–9.79	654–702	65.09–70.82	6.27–6.91	69060–89310	18040–18680	21580–25790
Post-salting	1100–1187	59.79–62.31	59.79–62.31	510–608	53.80–62.38	5.26–6.60	57710–77050	15720–18220	16950–23190
Drying	880–1110	52.27–54.63	52.27–54.63	422–585	42.39–48.03	5.66–7.70	58940–67210	12120–15980	12160–17090
Dry-cured hams	690–756	40.66–45.88	40.66–45.88	256–423	29.11–40.21	3.88–8.29	48580–62330	8120–11340	9860–14200

See abbreviations in Fig. 3.

tion retrieved from MRI–CVT (BV, SV and HV) procedures. Two methods in data mining were used, multiple linear regression and isotonic regression.

To validate the predicted results, the coefficient correlation R^2 of the two explored data mining methods was computed (Table 5). For weight, moisture content in B and S muscles and lipid content in B, high correlation coefficients (0.87–0.99) were obtained. Very few differences were found between correlation coefficients achieved by multiple linear regression and isotonic regression methods. The computational cost of both techniques is similar, and yet, isotonic regression is not automatic and needs a subsequent interpolating step by using a spreadsheet. Thus, the use of multiple linear regression for deducing these P-C parameters seems to be more comfortable.

Table 5
Correlation coefficient (R^2) for each physico-chemical characteristic predicted by applying data mining (multiple linear regression (MLR) and isotonic regression (IR)) on data achieved by MRI and Computer Vision Techniques (BV, SV and HV).

	BW	BM	BL	SW	SM	SL
MLR	0.954	0.966	0.871	0.937	0.969	0.035
IR	0.995	0.975	0.986	0.989	0.987	0.817

See abbreviations in Fig. 3.

In the case of lipid content in the S muscle, no good correlations were obtained when applying multiple linear regression, but accurate results were achieved ($R^2 = 0.817$) with isotonic regression. As previously explained, this could be related to the high variability of fat content in Iberian ham. In fact, the use of isotonic regression is indicated when having non-linear dependent data (Barlow et al., 1972).

Fig. 4 presents the adjustment between real and predicted values of lipid content in the S muscle by the two deductive techniques applied in this study. Isotonic regression shows higher accuracy in comparison to multiple linear regression for predicting the lipid content of S.

Table 6 shows prediction equations for weight, moisture and lipid content in the B and S muscles by multiple linear regression and isotonic regression. Thus, by using data obtained non-destructively by MRI–CVT (HV, BV and SV) weight, moisture and lipid content can be now reliable estimated. These determinations have always been carried out in Iberian hams, but the traditional methods are time-consuming and require the destruction of the sample. Therefore, our equations could be considered as a useful tool.

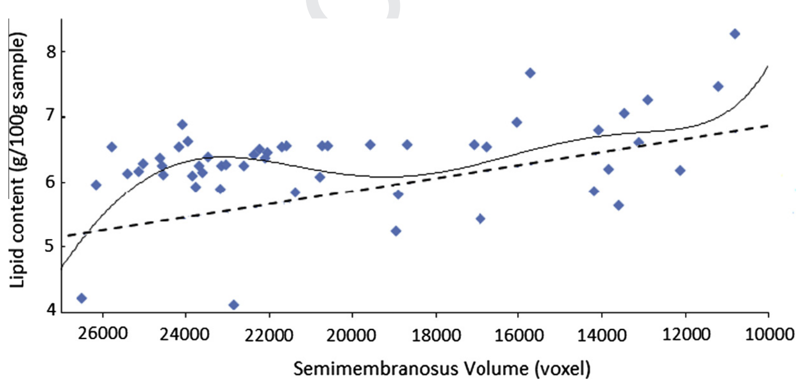

Fig. 4. Adjustment between real (♦) and predicted values of the lipid content of *Semimembranosus* muscle by using multiple linear regression (- - -) and isotonic regression (—) as a function of the *Semimembranosus* volume (expressed in voxel).

Table 6
Prediction equations of Iberian ham quality traits achieved by applying multiple linear regression (MLR) and isotonic regression (IR) on data achieved by MRI and Computer Vision Techniques (BV, SV and HV).

MLR	IR
$BW = 0.0445 * BV + 0.0131 * SV + 154.6591$	$BW = -4 \times 10^{-23} * HV^6 + 2 \times 10^{-17} * HV^5 - 3 \times 10^{-12} * HV^4 + 2 \times 10^{-7} * HV^3 - 0.0109 * HV^2 + 280.58 * HV - 3 \times 10^6$
$BM = 0.0021 * BV + 0.0002 * SV + 21.6042$	$BM = 2 \times 10^{-24} * SV^6 - 7 \times 10^{-19} * SV^5 + 5 \times 10^{-14} * SV^4 - 2 \times 10^{-9} * SV^3 + 3 \times 10^{-5} * SV^2 - 0.2673 * SV + 925.87$
$BL = -0.0007 * BV + 21.7736$	$BL = -2 \times 10^{-25} * HV^6 + 7 \times 10^{-20} * HV^5 - 1 \times 10^{-14} * HV^4 + 1 \times 10^{-9} * HV^3 - 8 \times 10^{-5} * HV^2 + 2.4017 * HV - 29387$
$SW = 0.0263 * BV + 0.0063 * SV + 28.4885$	$SW = 2 \times 10^{-21} * BV^6 - 3 \times 10^{-16} * BV^5 + 1 \times 10^{-11} * BV^4 - 3 \times 10^{-7} * BV^3 + 0.0044 * BV^2 - 29.538 * BV + 80337$
$SM = 0.0029 * BV + 0.0007 * SV - 4.2683$	$SM = 8 \times 10^{-23} * SV^6 - 1 \times 10^{-17} * SV^5 + 5 \times 10^{-13} * SV^4 - 1 \times 10^{-8} * SV^3 + 0.0002 * SV^2 - 1.2033 * SV + 3543.8$
$SL = -0.0001 * SV + 7.8575$	$SL = 2 \times 10^{-23} * SV^6 - 2 \times 10^{-18} * SV^5 + 1 \times 10^{-13} * SV^4 - 2 \times 10^{-8} * SV^3 + 3 \times 10^{-5} * SV^2 - 0.222 * SV + 636.56$

See abbreviations in Fig. 3.

4. Conclusions

To the best of our knowledge this work has been the first to apply data mining to Iberian ham information obtained from P-C analysis, weight, moisture and lipid content, and MRI-CVT techniques, volume.

The application of deductive techniques from data mining, multiple linear regression, to information from MRI-CVT and P-C analysis allows for the accurate estimation of more records of the analysed traits: weight, moisture content, lipid content, and volume in Iberian hams.

Multiple linear regression and isotonic regression are accurate methods of data mining for predicting weight, moisture and lipid content in Iberian ham as a function of features obtained from MRI-CVT techniques.

Data mining and MRI-CVT have been used as a pioneering approach to study the features of hams. These tools can be useful for calculating P-C parameters related to ham quality and for improving the control of the processing without destroying meat pieces.

5. Uncited reference

Q4 Reutermann (2012).

Acknowledgments

The authors wish to acknowledge the funding received for this research from both the Junta de Extremadura (Regional Government Board – Research Projects 3PR05B027 and PDT08A021; Consejería de Economía, Comercio e Innovación and FEDER – economic support for research groups: GRU09148 and GRU09025) and from the Spanish Government (National Research Plan) and the European Union (FEDER funds) by means of the grant reference TIN2008-03063. We also wish to thank the "Hermanos Roa" company from Villar del Rey (Badajoz), as well as the "Infanta Cristina" University Hospital Radiology Service, specially to the Dr. Ramón Palacios, for their contribution and support.

References

Andres, A.I., Ventanas, S., Ventanas, J., Cava, R., Ruiz, J., 2005. Physicochemical changes throughout the ripening of dry cured hams with different salt content and processing conditions. Eur. Food Res. Technol. 221, 30–35.
Antequera, T., Caro, A., Rodriguez, P.G., Pérez-Palacios, T., 2007. Monitoring the ripening process of Iberian Ham by computer vision on magnetic resonance imaging. Meat Sci. 76, 561–567.
Association of Official Analytical Chemist (AOAC), 2000. Official Methods of Analysis of the Association of Official Analytical Chemists, seventeenth ed. Gaithersburg, Maryland.
Ávila, M., Durán, M.L., Caro, A., Antequera, T., Gallardo, R., 2005. Thresholding methods on MRI to evaluate iIntramuscular fat level on Iberian ham. Lectures Notes in Computer Science (LNCS 3523). Pattern Recognition and Image Analysis, 697–704.
Barlow, R.E., Bartholomew, D., Bremner, J.M., Brunk, H.D., 1972. Statistical Inferece Under Order Restriction: The Theory and Application of Isotonic Regression. Wiley, New York.
Borge, L., 1985. Estimacion y contrastes de hipótesis en el modelo lineal general con restricciones de desigualdad. Doctoral thesis. University of Valladolid, Spain.
Caro, A., Rodríguez, P.G., Cernadas, E., Durán, M.L., Villa, D., 2001. Applying active contours to muscle recognition in Iberian ham MRI. In: IASTED International Conference Signal Processing, Pattern Recognition and Applications, Rhodes, Greece.
Caro, A., Durán, M.L., Rodríguez, P., Antequera, T., Palacios, R., 2003. Mathematical morphology on MRI for the determination of Iberian ham fat content. Lecture Notes in Computer Science (LNCS 2905). Prog. Pattern Recogn. Speech Image Anal., 359–366.
Collell, C., Gou, P., Arnau, J., Comaposada, J., 2011. Non-destructive estimation of moisture, water activity and NaCl at ham surface during resting and drying using NIR spectroscopy. Food Chem. 129, 601–607.
Cortez, P., Portelinha, S., Rodrigues, S., Cadavez, V., Teixeira, A., 2006. Lamb meat quality assessment by support vector machines. Neural Process. Lett. 24, 41–51.
Cortez, P., Cedeira, A., Almeida, F., Matos, T., Reis, J., 2009. Modeling wine preferences by data mining from physicochemical properties. Decis. Support Syst. 47, 547–553.
Fantazinni, P., Gombia, M., Schembri, P., Simoncini, N., Virgili, R., 2009. Use of magnetic resonance imaging for monitoring Parma dry-cured ham processing. Meat Sci. 82, 219–227.
Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. Al Mag. 17, 37–54.
Fulladosa, E., Santos-Garcés, E., Picouet, P., Gou, P., 2010. Prediction of salt and water content in dry-cured hams by computed tomography. J. Food Eng. 96, 80–85.
Furnols, M.F., Teran, M.F., Gispert, M., 2009. Estimation of lean meat content in pig carcasses using X-ray Computed Tomography and PLS regression. Chemometr. Intell. Lab. Syst. 98, 31–37.
Haseth, T.T., Sørheim, O., Høy, M., Egelandsdal, B., 2012. Use of computed tomography to study raw ham properties and predict salt content and distribution during dry-cured ham production. Meat Sci. 90, 858–864.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining. Inference and Prediction Springer-Verlag, New York.
Holmes, G., Fletcher, D., Hernández, J., Ramírez, M.J., Ferri, C., 2007. Introducción a la Minería de Datos. Prentice-Hall.
Klaypradith, W., Kerdpiboon, S., Singh, R.K., 2010. Application of artificial neural networks to predict the oxidation of menhaden fish oil obtained from Fourier transform infrared spectroscopy method. Food Bioprocess Technol. 4, 475–480.
Manzoco, L., Anese, M., Marzona, S., Innocente, N., Lagazio, C., Nicoli, M.C., 2013. Monitoring dry-curing of S. Daniele ham by magnetic resonance imaging. Food Chem. 141, 2246–2252.
Martin, L., Córdoba, J.J., Antequera, T., Timon, M.L., Ventanas, J., 1998. Effects of salt and temperature on proteolysis during ripening of Iberian ham. Meat Sci. 49, 145–153.
Mitchell, T.M., 1999. Machine learning and data mining. Commun. ACM 42, 30–36.
Pérez-Juan, M., Afseth, N.K., González, J., Díaz, I., Gispert, M., Furnols, M.F., Oliver, M.A., Realini, C.E., 2010. Prediction of fatty acid composition using a NIRS fibre optics probe at two different locations of ham subcutaneous fat. Food Res. Int. 43, 1416–1422.
Pérez-Palacios, T., Ruiz, R., Martin, D., Muriel, E., Antequera, T., 2008a. Comparison of different methods for total lipid quantification. Food Chem. 110, 1025–1029.
Pérez-Palacios, T., Ruiz, J., Antequera, T., 2008b. Perfil de ácidos grasos de la grasa subcutánea e intramuscular de credos ibéricos cebados en montanera y con pienso "alto oleico". Eurocarne, 1–10.
Pérez-Palacios, T., Antequera, T., Durán, M.L., Caro, A., Rodríguez, P.G., Ruiz, J., 2010a. MRI-based analysis, lipid composition and sensory traits for studying Iberian dry-cured hams from pigs fed with different diets. Food Chem. 126, 1366–1372.
Pérez-Palacios, T., Antequera, T., Molano, R., Rodríguez, P.G., Palacios, R., 2010b. Sensory traits prediction in dry-cured hams from fresh product via MRI and lipid composition. J. Food Eng. 101, 152–157.
Pérez-Palacios, T., Ruiz, J., Dewettinck, K., Trung Le., T., Antequera, T., 2010c. Individual phospholipid classes from Iberian pig meat as affected by diet. J. Agri. Food Chem. 58, 1755–1760.
Pérez-Palacios, T., Ruiz, J., Martín, D., Barat, J.M., Antequera, T., 2011a. Pre-cure freezing effect on physicochemical, texture and sensory characteristics of Iberian ham. Food Sci. Technol. Int. 17, 127–133.
Pérez-Palacios, T., Antequera, T., Durán, M.L., Caro, A., Rodríguez, P.G., Palacios, R., 2011b. MRI-based analysis of feeding background effect on fresh Iberian ham. Food Res. Int. 43, 248–254.
Picouet, P.A., Gou, P., Fulladosa, E., Santos-Garcés, E., Arnau, J., 2013. Estimation of NaCl diffusivity by computed tomography in the Semimembranosus muscle during salting of fresh and frozen/thawed hams. LWT – Food Sci. Technol. 51, 275–280.
Reutermann, P., 2012. An application of data mining to fruit and vegetable sample identification using gas chromatography-mass spectrometry. In: Proceedings of International Congress of Enviromental Modelling and Software Managing Resources of a Limited Planet, Leipzig, Germany.
Ruiz, J., García, C., Muriel, E., Andrés, A.I., Ventanas, J., 2002. Influence of sensory characteristics on the acceptability of dry-cured ham. Meat Sci. 61, 347–354.
Santos-Garcés, E., Gou, P., Garcia-Gil, N., Arnau, J., Fulladosa, E., 2010. Non-destructive analysis of aw, salt and water in dry-cured hams during drying process by means of computed tomography. J. Food Eng. 101, 187–192.
Song, Y.H., Kim, S.J., Lee, S.K., 2002. Evaluation of ultrasound for prediction of carcass meat yield and meat quality in Korean native cattle. Asian J. Animal Sci. 15, 591–595.
Vestergaard, C., Erbou, S.G., Thauland, T., Adler-Nissen, J., Berg, B., 2005. Salt distribution in dry-cured ham measured by computed tomography and image analysis. Meat Sci. 69, 9–15.
Witten, I.H., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, San Francisco.
Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. Knowl. Inf. Syst. 14, 1–37.