Artificial intelligence techniques point out differences in classification performance between light and standard bovine carcasses

J. Díez[a], A. Bahamonde[b], J. Alonso[b], S. López[b], J.J. del Coz[b], J.R. Quevedo[b], J. Ranilla[b], O. Luaces[b], I. Alvarez[a], L.J. Royo[a], F. Goyache[a],*

[a]*SERIDA-CENSYRA-Somió, C/Camino de los Claveles 604, E-33203 Gijón (Asturias), Spain*
[b]*Centro de Inteligencia Artificial. Universidad de Oviedo at Gijón, Campus de Viesques, E-33271 Gijón (Asturias), Spain*

Abstract

The validity of the official SEUROP bovine carcass classification to grade light carcasses by means of three well reputed Artificial Intelligence algorithms has been tested to assess possible differences in the behavior of the classifiers in affecting the repeatability of grading. We used two training sets consisting of 65 and 162 examples respectively of light and standard carcass classifications, including up to 28 different attributes describing carcass conformation. We found that the behavior of the classifiers is different when they are dealing with a light or a standard carcass. Classifiers follow SEUROP rules more rigorously when they grade standard carcasses using attributes characterizing carcass profiles and muscular development. However, when they grade light carcasses, they include attributes characterizing body size or skeletal development. A reconsideration of the SEUROP classification system for light carcasses may be recommended to clarify and standardize this specific beef market in the European Union. In addition, since conformation of light and standard carcasses can be considered different traits, this could affect sire evaluation programs to improve carcass conformation scores from data from markets presenting a great variety of ages and weights of slaughtered animals.
© 2003 Elsevier Science Ltd. All rights reserved.

Keywords: Bovine carcass; Conformation assessment; SEUROP; Artificial intelligence; Machine learning; Relevancy

1. Introduction

Carcass conformation assessment is aimed at facilitating the relationships between market operators and supplying markets with quality products to satisfy consumer demands. Since the CEE 390/81, CEE 1208/81, CEE 2930/81 and CEE 1026/91 regulations, the European Union has set up a standard bovine carcass conformation assessment system, known as the SEUROP system, to be applied throughout the EU. This grading system is expected to be useful for bovines from 300 Kg live weight, regardless of sex, age or management conditions of the animal. To be applicable for such a range of products, the SEUROP system is described very broadly. The description of conformation classes ranges from 'exceptional' to 'poor' muscular development and from 'extremely' convex to 'very' concave profiles.

In applying such a broad classification system, it is not surprising that classifiers need a complex training procedure before carrying out commercial classifications in order to avoid substantial individual differences in assessments. However, the repeatability of grading tends to be low, thus affecting the market's confidence. Carcass classification by line-graders can give errors that can reach 15% (George et al., 1996). This situation is more evident in markets such as Northern Spain, Portugal and in Italy, where there exists a significant proportion of light carcasses. Usually, it is accepted that the conformation score tends to increase with age and carcass weight (Kempster, Cook, & Southgate, 1988; More O'Ferral, & Keane, 1990). However, most studies have been made with carcasses aged between 14 and 24 months. In this age range, the animals have reached a sufficient degree of physiological maturity to be considered young adults, but light carcasses come from animals showing only the earlier stages of puberty. The shapes of the body and carcasses of these young animals cannot be easily compared with those of physiologically mature animals. Thus, since carcasses

---

* Corresponding author. Tel.: +34-985-19-53-03; fax: +34-985-19-53-10.

E-mail address: fgoyache@serida.org (F. Goyache).

are graded by comparison with the standard carcass model, the application of the SEUROP system to light carcasses can be problematic.

To avoid errors in carcass grading, there have been attempts to develop automatic classifiers of carcass conformation (Belk, Scanga, Tatum, Wise, & Smith, 1998; Cannell et al., 1999; Goyache, Bahamonde et al., 2001) using analysis of digital images. However, assessments carried out by expert line-graders continue to be essential for carcass classification in practical terms. In fact, there are no standard instruments to replace human graders, in part because these instruments would be required to predict both carcass composition and palatability of the cooked product (Smith, 1999).

The aim of this paper is not to propose an automatic assessment system to replace the human grader. We are concerned with testing the validity of the assumptions of the official SEUROP conformation classification system for light carcasses, identifying, by means of Artificial Intelligence (AI) techniques, the major factors affecting conformation grading of light compared with standard carcasses.

We use AI systems to synthesize the knowledge required to accomplish useful tasks from a sample of intelligent behavior; in other words, we use Machine Learning (ML) algorithms. This methodology has been successfully applied in animal production and the food industry (Goyache, del Coz et al., 2001; Goyache, Bahamonde et al., 2001). Expert classifiers perform the SEUROP system by means of a kind of knowledge that can be computationally represented. Recently, we have shown the possibility of using AI techniques to clone the behavior of bovine carcass classifiers to obtain sound assessments (Goyache, Bahamonde et al., 2001). But AI can be used more broadly than the production of computer routines to be used in an industrial environment. We will use the ability of AI algorithms to explain the knowledge learned, in order to gain an insight into the major features affecting the performance of the present classification system for bovine carcass.

2. Materials and methods

2.1. Data

The input data for learning algorithms are called training sets, i.e. collections of examples of the behavior that we would like to learn to repeat. In our case, each example is described by means of attributes that numerically represent the traits of the carcasses considered. Additionally, the examples are given the conformation score assigned by one of our experts; we will refer to this score as the *class* of the example.

Our training set is a representative subset of the examples used previously (Goyache, Bahamonde et al., 2001), and includes the hot carcass weight and conformation score from a total of 97 bovine carcasses. Carcasses were graded individually, following the SEUROP system, by three expert classifiers of EASA, the control organization of the Quality Beef Program "*Carne de Asturias*". These carcasses were selected to have a representative sample of the actual distribution of sexes, conformation classes and weights existing in the entire beef market in Asturias (Northern Spain). It should be noted that most of the beef production in this region is based on well-conformed "*Asturiana de los Valles*" animals (Goyache, Bahamonde et al., 2001). To improve the accuracy of classification, the signs $+$ or $-$ could be added to each conformation grade (Kempster, 1986). The SEUROP grades were numerically codified from 1 (P) to 6 (S); the signs $+$ and $-$ were computed as $+0.25$ or -0.25 conformation points, respectively (Goyache, Bahamonde et al., 2001). Each classifier graded most carcasses: the first one classified 78 carcasses, the second 74, and the third 75 giving 227 classification events, which make up the training set as independent training examples.

To capture carcass traits, our experts photographed each carcass in three different positions: (a) lateral, (b) medial and (c) dorsal. The digital pictures were then processed by marking (with mouse clicks) 21 key points and five curve arcs (see Fig. 1) to calculate up to 28 attribute values numerically representing the features or traits mentioned in SEUROP specifications. This methodology was successfully used previously (Goyache, del Coz et al., 2001; Goyache, Bahamonde et al., 2001). The attributes used are listed in Table 1. Our techniques are not only useful for assessing carcass dimensions, as occurs with traditional methodologies (De Boer, Dumont, Pomeroy, & Weniger, 1974), but also for estimating differences in muscular development and carcass profiles. We included a metric reference in each photo so as to be able to measure lengths from the pictures. Single anatomical traits were easily calculated by means of the distance between key points (Fig. 2). Ratios and volumes (to estimate muscular development) were calculated combining some single anatomical traits. Representation of profiles was obtained by fitting, using minimal least squares, a 4th degree polynomial to each profile EC1, EC2, LC1, LC2, LC3. Then, the curvature at a point is computed from the first and the second derivative, as stated in Formula 1 of Fig. 2. Finally, we summarize the convexity of the profile by calculating the average of the curvature at a sequence of equidistant points placed over the arc that follows the profile (see Figs. 1 and 2 and Goyache, Bahamonde et al., 2001, for a more detailed description of the computation).

The blockiness index, calculated as the ratio between carcass weight and carcass length was included in the training set. Despite this trait not being properly included in the SEUROP rules, it was considered in our study because this attribute has been found to be rele-
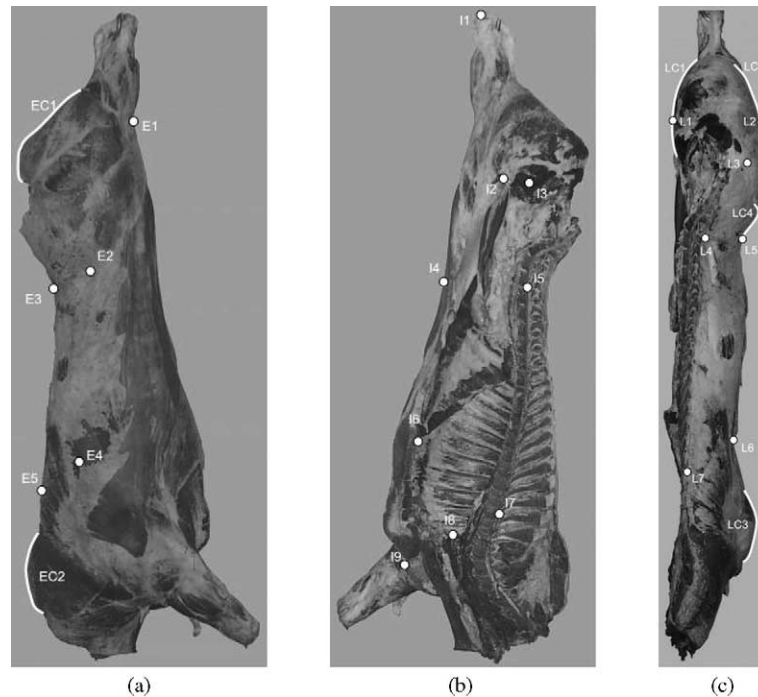
Fig. 1. Each carcass was photographed in three positions: lateral (a), medial (b) and dorsal (c). An operator then manually marked 21 relevant points and five curve arcs. Single anatomical traits were easily calculated by means of distances between key points (i.e. belly depth = distance between I4 and I5 or carcass length = distance between I2 and I7 in picture b). To represent profile convexities, we consider the curve arc that borders the profile as a variable real function f. We can then compute the curvature at each point $(x, y = f(x))$ by means of Formula 1. We approximate the derivates using the values of f in the environment of each point of the profile. So we divide the arc by means of a sequence of points $\{x_i\}$ in $[0,a]$ that divide the interval into a given number (the same in all cases) of equal length subintervals. Then $f'(x_i)$ and $f''(x_i)$, the first and second derivative, are approached using Formula 2 and Formula 3. Finally, to summarize the convexity of the whole arc in the interval $[0,a]$, we compute the average of the curvature (x_i) for all $\{x_i\}$.

vant to give an accurate carcass classification (Goyache, Bahamonde et al., 2001; Vallejo et al., 1992). However, to differentiate the influence of the blockiness index on ML algorithms, all the computations were carried out using two different initial training sets: with and without this attribute to describe the carcasses. Thus, the initial training sets respectively included 28 and 27 attributes describing carcass conformation.

Each initial training set was split into two subsets for light and standard carcasses. Light carcasses were considered those weighing less than 220 kg, following the commercial practice carried out in the most important Spanish beef markets, such as Binefar and Lleida. These markets are considered as a reference for the whole Spanish beef market. The light carcass subset included data from 29 carcasses and 65 training examples, while the standard carcass subset included data from 68 carcasses and 162 training examples. Descriptive statistics of the attributes defining our training sets are listed in Table 1. Each training subset, including the blockiness index or not, was used as an independent input to ML algorithms.

2.2. Machine learning algorithms

Different kinds of ML algorithms may perform differently in a given situation. Thus, we need to use several

ML algorithms to evaluate both the difficulty of the learning process and the usefulness of the acquired knowledge. At the same time, the representation of the induced knowledge may be quite different, and so we can appreciate new aspects of a given classification problem when we use ML algorithms of different families. A detailed description of AI principles and methods and of the main characteristics of several kinds of ML algorithms can be found in Goyache, del Coz et al. (2001) and Goyache, Bahamonde et al. (2001) and in the references cited therein.

Sometimes, approaches to applying ML procedures to carcass and meat research employ Artificial Neural Networks (ANNs) (Hill, Jones, Robertson, & Major, 2000). Unfortunately, ANNs need a previous definition of the layout and other parameters prior to performing the learning process. Thus, we cannot always know with certainty whether a possible failure in the learning process is a consequence of deficiencies in the customization process or in the learning ability of training sets. For this reason, we did not use learning algorithms based on ANNs. Thus, we are concerned with the use of ML algorithms able to make accurate predictors with only the input of the training sets.

The topics highlighted above led us to use the following ML algorithms in the present analysis:

Table 1
Description of the attributes used to represent bovine carcass conformation grading in 25 carcasses building the light carcasses training set and in 82 carcasses building the standard carcasses training set

	Light carcasses		Standard carcasses	
	Mean	SD	Mean	SD
Carcass weight*	188.52	24.86	290.27	64.34
Thigh width*	26.54	3.41	30.28	7.64
Thigh length*	43.44	3.025	46.88	3.05
Thigh ratio (width/length)	0.61	0.067	0.64	0.18
Hind limb length*	64.12	4.71	66.95	6.68
Shoulder length*	40.01	4.63	42.72	4.93
Shoulder width	33.40	3.52	38.06	11.98
Lateral shoulder height*	3.74	1.35	5.33	2.04
Shoulder area*	1342.00	243.20	1609.00	575.40
Shoulder volume*	5151.00	2460.00	9462.00	10,911.00
Round profile	−0.020	0.008	−0.019	0.009
Shoulder profile	−0.014	0.010	−0.017	0.010
Topside profile	−0.025	0.008	−0.023	0.011
Back length*	53.08	10.84	60.43	17.64
Front width of back*	10.72	4.33	12.48	3.90
Rear width of back	11.22	3.85	13.69	4.38
Back width ratio (front/rear)	0.93	0.26	0.90	0.30
Lateral back width at hip*	12.32	2.61	14.60	2.43
Lateral back width at chest*	13.98	2.32	15.43	2.80
Lateral back width ratio(hip/chest)	0.90	0.21	0.95	0.18
Back volume	7527.00	6384.00	13,344.00	18,182.00
Hips profile	−0.026	0.058	−0.032	0.046
Thigh profile	−0.023	0.009	−0.027	0.023
Neck profile	0.33	1.72	−0.016	0.013
Carcass length*	108.95	11.43	116.73	14.10
Chest depth*	41.59	4.71	44.96	5.99
Belly depth*	31.50	5.13	34.20	5.55
Blockiness index*	1.74	0.22	2.51	0.55
Conformation score	3.54	1.12	3.90	1.14

Attributes are described in cm for distances between two given key points, cm^2 for areas and cm^3 for volumes; profiles and ratios are described in the corresponding units; carcass weight is in kg. The symbol * as a superscript expresses significantly different means ($P < 0.05$). Conformation score was obtained following SEUROP methodology; classifiers may added signs + or −to each conformation grade; SEUROP grades were numerically codified ranging from 1 (P) to 6 (S); the signs + and −were computed as +0.25 or −0.25 conformation points, respectively.

- Cubist (2000). This is a ML algorithm inducing a set of regression rules, i.e. constructs of the form:

  if   some conditions about the traits are fulfilled
  then to compute the SEUROP conformation score use a given linear function

- SAFE (System to Acquire Functions from Examples) (Quevedo & Bahamonde, 1999). This ML algorithm also returns a set of regression rules, like Cubist. However, the conditions of their rules do not cover all the possible values of the attributes, so when they are asked to classify a case that does not fulfil any of the rule conditions, the predicted score for that case will be computed by interpolating (inversely to their distance to the case) the values provided by the nearest rules. This evaluation procedure is usually called partial matching or fuzzy evaluation.

- BETS (Best Examples in Training Sets) (Bahamonde, de la Cal, Ranilla, Alonso, J., 1997; Del Coz et al., 1999; Luaces et al., 1999) is a ML algorithm of another family. It selects the most representative instances of the training set and considers that these examples are useful for classifying the

$$k(x_i) = \frac{f''(x_i)}{\left(1 + f'(x_i)^2\right)^{3/2}} \qquad K = \frac{\sum_{i=1}^{n} k(x_i)}{n}$$

| Formula 1 | Formula 2 |

Fig. 2. Formulas used to compute the curvature of a profile described by an arc of a curve $y = f(x)$. The sequence (x_i, $i = 1, …, n$) are equidistant points in the x axis of the arc considered.

elements of the domain. To predict the numeric value of an unseen case *q*, BETS searches for the two nearest remembered instances and then returns a score inversely related to the distance from *q* of the values attached to the nearest neighbors. From a conceptual point of view, BETS induces regression rules where linear functions are always constant, and conditions are defined by intervals with only one point. In straightforward terms, BETS' regression rules are simple examples selected from the training set where the attribute values considered irrelevant are omitted.

The accuracy of the performance of each ML algorithm was estimated through a *cross validation* system. Each training set was divided into 10 folders. Each of these folders was successively used as a test set while the other nine were used for training. The prediction function obtained by the ML algorithm from the other nine folders was then applied to each example from the test folder, and then the average absolute difference with respect to the class of the example was computed. The experiment was run 5 times, finally returning the average of the differences thus computed. This procedure gives a good estimation of the accuracy of the classification function learned from the whole training set when we apply it to unseen cases. This particular feature is very useful when discussing the results of the present study.

Another important issue in Machine Learning is the number of functions or prototypes (examples remembered for future classifications) needed to explain the computer assessments. When we use learners that return sets of regression rules, the number of such rules is a good measure of the linearity of the assessing task. In the case of BETS, the number of prototypes selected with respect to the total is also reliable evidence about how non-linear the assessing task is.

In addition to the ML algorithms mentioned above, we used classical linear regression to make sure that the accuracy of the results from the carcass classification would not be affected by bias due to the learning method.

2.3. Relevancy

Frequently, we do not know what the most important attributes are in order to solve problems concerning knowledge. In fact, ML techniques are often used when the factors affecting a process are not well known. This is the case with the classification of light or standard bovine carcasses. We can describe bovine carcasses using a broad list of attributes. Of course, not all the attributes have the same weighting to grade SEUROP conformations. Sometimes their contribution is only redundant or negligible information. It is clear that if it

were possible to ascertain the most discriminate attributes affecting a problem, its solution would not only be easier, but also more accurate and generally applicable.

Thus, when we are trying to induce knowledge, the study of the relevancy of the attributes that describe the examples is a core issue. Many ML algorithms include some type of mechanism to select the major attributes affecting the studied classification issue. In all cases, the underlying idea is to be able to estimate the prediction quality of the attribute or attribute values in order to make a classification decision. Thus, the algorithms used to study relevancy are usually based on learners used to estimate the accuracy in different attribute arrangement assumptions; see, for instance (Blum & Langley, 1997; Kira & Rendell, 1992; Wettschereck, Aha, & Mohri, 1997).

In our study, the training sets initially included 28 attributes; 27 if we exclude the blockiness index of the carcasses. As stated above, our aim was to find the best combination of attributes presenting the best ratio between the prediction error in what was learned and the number of attributes used to learn. This search is very costly in computational terms and is carried out following a kind of heuristic search.

To accomplish this stage, a combination of tools successfully implemented previously (Goyache, Bahamonde et al., 2001) was used to estimate the relevancy of the attributes involved in the description of the carcasses. So, as a side effect of BETS' learning process, we can numerically obtain the degree to which the value of a given attribute can help to identify a given class. Then it is possible, using these results, to obtain an ordering of the relevancy of the attributes. A differentiation was made with this ordering between relevant and irrelevant attributes. The less informative attributes should then be removed. To do so, a second tool called FA (Goyache, Bahamonde et al., 2001; Quevedo, del Coz, & Díez, 2001) was applied. This tool removes the less relevant attribute and, in a subsequent step, checks the usefulness of the obtained training set (smooth screening). This check is made by a cross-validation methodology using the nearest-neighbor algorithm (the number of neighbors to be used is calculated as a function of the number of examples). This process is repeated iteratively, removing the attributes one by one, with the exception of the most relevant one. FA will select the training set made by the combination of attributes producing the least error. If the least error were obtained using the initial training set, this would be selected. Sometimes, however, the selected training set is still too complex to be easily managed, so FA can carry out a more rigorous screening of the attributes, accepting in this case a small increase in the error obtained by the resultant training set after removing the less relevant attributes (hard screening). The permitted increase in error can be a user-defined parameter; in this

paper we allowed a maximum error of 0.50, which is the numerical gap between the signs $+$ and $-$ for each conformation grade.

3. Results

Table 1 presents descriptive statistics of the training sets of light and standard carcasses. Average carcass weight was 188.5 kg for light carcasses, and 290.3 kg for standard carcasses. Standard carcasses were approximately 8 cm longer and 3.5 cm deeper at the chest than light carcasses and back and thigh lengths were about 7 and 3.5 cm greater for standard carcasses. Attributes describing carcass size and body measurements present significant differences between the carcass types. However, it should be noted that the means of the conformation score and the curvature of profiles were not significantly different between light and standard carcasses. Major attributes in carcass muscular development description, such as thigh ratio, shoulder width or back volume, did not present significant differences. We additionally found that standard carcasses were significantly more compact than light carcasses ($+0.77$ units).

Table 2 shows the results of the learning process using three ML algorithms, together with the results obtained using classical linear regression. Average absolute errors were found by means of cross validation; this guarantees that they are sound estimations of the absolute value of the difference in carcass conformation scores between the predicted value and the average of our human experts. We thus found that the results obtained by means of linear regression are higher than those obtained using ML algorithms, especially when hard screening substantially reduces the number of attributes used as input to classify carcass conformation. Regardless of the inclusion or not of the blockiness index in the training sets, errors obtained with the lower number of attributes used to classify light carcasses are lower than those obtained for standard carcasses under the same conditions, ranging between 0.365 and 0.447 and between 0.436 and 0.468, respectively. BETS obtained the best results for light carcasses (0.365) and SAFE for standard carcasses (0.445) when we include the blockiness index in the training set. When this attribute is not taken into account, BETS and CUBIST gave similar results for both light (0.40) and standard carcasses (around 0.44).

Table 2
Average absolute errors and number of functions (or prototypes) selected to make predictions on unseen cases for three machine learning algorithms in light or standard carcass conformation classification. Results obtained by cross validation of 10 folders repeated five times over the light and standard carcasses training sets described in the text. The errors were computed as the deviation from the human experts' average assessment

Number of attributes used	Linear regression	BETS		SAFE		Cubist	
	Error	Error	Prototypes	Error	Functions	Error	Functions
With blockiness index							
Light carcasses subset							
28	0.58	0.38	9.76	0.37	1.26	0.43	4.08
12[a]	0.34	0.32	10.80	0.34	1.80	0.39	1.80
11[b]	0.34	0.32	10.44	0.36	1.84	0.37	1.82
4[c]	0.50	0.37	9.50	0.40	4.34	0.41	4.58
Standard carcasses subset							
28	0.45	0.44	24.98	0.42	3.28	0.42	5.16
9[a]	0.49	0.39	24.84	0.42	7.00	0.46	4.20
8[b]	0.49	0.40	25.00	0.43	6.28	0.46	3.64
6[c]	0.49	0.39	24.42	0.42	6.74	0.40	5.88
3[c]	0.56	0.47	23.66	0.45	8.06	0.46	6.46
Without blockiness index							
Light carcasses subset							
27	0.58	0.38	9.58	0.37	1.34	0.40	3.54
11[a]	0.34	0.35	10.30	0.34	1.72	0.37	2.22
10[b]	0.33	0.32	11.04	0.35	1.96	0.36	1.92
3[c]	0.57	0.40	9.40	0.45	4.32	0.40	4.58
Standard carcasses subset							
27	0.45	0.44	24.06	0.43	3.06	0.42	5.58
8[a]	0.51	0.42	22.60	0.47	5.92	0.45	5.52
7[b]	0.53	0.44	23.78	0.46	6.54	0.44	6.00

[a] Smooth screening.
[b] Hard screening.
[c] Subsequent harder screenings.

SAFE and CUBIST need around two linear functions when they classify light carcasses with a large number of attributes, but when the number of attributes involved is reduced, the learned classification mechanism becomes non-linear, using at least four regression rules. The number of linear functions used to classify standard carcasses is slightly higher, ranging from 3 to 8, depending on the number of attributes used. On the other hand, in all cases BETS selects around the 16% of the training instances to be used as models for classifying unseen cases.

The application of techniques to assess the relevance of the attributes on carcass classification has quite a different influence on light or standard carcass subsets. The application of smooth screening using FA (Quevedo et al., 2001) produces a dramatic reduction (60%) in the number of attributes used to learn how to classify carcasses. However, it is necessary to carry out at least two consecutive hard screenings to obtain a substantial reduction in the number of attributes used without exceeding the maximum error admitted in our analysis (0.50). Thus, there are a large number of attributes that are taken into account or rejected together. This is more evident for standard carcasses: harder screening, in two consecutive steps, succeeds in reducing the attributes from 8 to 3 when the blockiness index is taken into account; while when this attribute is not considered, we must use seven attributes to obtain a sound classification. To reduce the number of attributes, we would have to allow average absolute errors in classification performance higher than 0.50.

The smallest number of attributes used to learn how to classify bovine carcasses in each training subset are listed in Table 3. The blockiness index seems to be very informative, regardless of the type of carcass; when this attribute is included, it is always selected as relevant. Its inclusion enables a reduction in the number of attributes required to obtain sound assessments of standard carcasses. To classify light carcasses, our

classifiers considered muscular development (thigh width and thigh ratio) and body development (belly depth). To classify standard carcasses, however, classifiers basically use attributes characterizing muscular development (shoulder width) and muscular profiles (topside profile). The other attributes used to classify standard carcasses affect muscular development of the back and, to a lesser extent, of the thigh (thigh length).

4. Discussion

The measurements and weights shown in Table 1 are typical for the beef market in Asturias (Northern Spain). The significant differences estimated between attributes describing carcass size, such as carcass weight, carcass length or thigh and back measurements, are not surprising. The threshold we used to build our different subsets, based on the accepted practice of major Spanish market operators, reflects an actual break in continuity between light and standard carcasses in the analyzed database. The lack of significant differences between curvature of profiles and conformation scores for both types of carcasses may be more surprising. The conformation score is expected to increase with age and carcass weight (Kempster et al., 1988; More O'Ferral & Keane, 1990) during fattening, as conformation is a visual assessment of the thickness of muscle and fat in relation to the size of the skeleton. The lack of significant differences in the present study may be explained by the origin of the animals. The beef market in Asturias consists mainly of pure bred *Asturiana de los Valles* animals. This breed shows a high proportion of double-muscled animals producing well-conformed carcasses regardless of the age of the animals (Goyache, Royo, Alvarez, & Gutiérrez, 2002; Gutiérrez & Goyache, 2002). However, our results highlight the possibility that classifiers give the same score to different products. Of course, a general impression of the whole

Table 3
List of attributes used to learn how to classify light or standard carcasses for each of the subsets considered in the present paper after assessing their relevancy by means of the FA tool (Quevedo et al., 2001). The list of attributes is ordered from less to more relevant. Attributes have been listed when there were seven or less of them

Number of attributes	Attributes used
With blockiness index	
Light carcasses subset	
4	Belly depth, Blockiness index, Thigh width, Thigh ratio
Standard carcasses subset	
6	Rear width of back, Back length, Back width ratio, Topside profile, Blockiness index, Shoulder width
3	Topside profile, Blockiness index, Shoulder width
Without blockiness index	
Light carcasses subset	
3	Belly depth, Thigh width, Thigh ratio
Standard carcasses subset	
7	Thigh length, Shoulder profile, Rear width of back, Back length, Back width ratio, Topside profile, Shoulder width

carcass affects the grading. This is reflected by the relevancy of the blockiness index in the training sets. However, the behavior of classifiers seems to be quite different when they grade a light or a standard carcass. Classifiers apply SEUROP rules more rigorously when they classify standard carcasses. For this type of carcass, our experts use only attributes characterizing muscular development in shoulder, back and thigh and carcass profiles. In this sense, despite the importance given by the SEUROP rules to profile assessment, only two of them (especially topside profile) are taken into account in the classifiers decisions. Moreover, most muscular development attributes, especially back attributes, have a combined influence; they are taken into account or rejected altogether (see Table 3). In addition, carcass profiles seem to be of little importance when classifying light carcasses. To reach accurate light carcass classifications, our experts do not match the SEUROP rules exactly. To do this, they use attributes characterizing muscular and body development.

The present results may be compared with those we have recently reported (Díez et al., 2001; Goyache, Bahamonde et al., 2001), regardless of the type of carcass. In these previous papers, we reported average absolute errors of four ML algorithms ranging between 0.45 and 0.49, while the average absolute errors of each expert with respect to the average SEUROP scores of the three experts were 0.41 for the first, 0.39 for the second and 0.41 for the third. The errors obtained in the present study are, in general, similar or slightly lower than those reported previously when classifying standard carcasses, and substantially lower than reported for light carcass classification. Despite the smaller number of examples, the light carcass subset is more homogeneous than the standard carcass subset, leading to a better performance of ML algorithms.

In addition, the behavior of classifiers seems to be quite different when dealing with a light or a standard carcass. We reported previously (Díez et al., 2001; Goyache, Bahamonde et al., 2001) that carcass classification presents highly non-linear behavior. Now we know that this is true irrespective of the type of carcass. However, the attributes used by our ML algorithms in the present analysis are quite different for each carcass type and contrast with those assessed as relevant for grading carcass conformation regardless of its type. If we try to classify carcasses over the whole available range, we should use four (belly depth, blockiness index and round and topside profiles) or five attributes (chest depth, carcass length, topside profile, round profile and belly depth), respectively, when the blockiness index is included (Goyache, Bahamonde et al., 2001) or not (Díez et al., 2001) in the training set. Thus when classifying the whole set of carcasses and regardless of whether we include the blockiness index, we can say that the attributes characterizing body size or skeletal develop-

ment, such as chest depth, carcass length, and belly depth, acquire considerable importance overriding the attributes characterizing muscular development, while major profile attributes remain important.

The present results confirm the different underlying behavior of conformation classifiers when dealing with a given type of carcass. In fact, they consider different attributes when grading different carcass classes under SEUROP assumptions. The estimated differences might, however, be considered to be a consequence of sample error. To eliminate this possibility, we got the rules induced by Cubist to classify light or standard carcasses using the training subsets including the lowest number of attributes when we use the blockiness index. We then used the rules obtained in a given carcass subset to reassess the examples of the other subset. The error in standard carcass classification by means of rules learned to classify light carcasses was 0.97. When we classify light carcasses by means of the rules obtained when classifying standard carcasses using three or six attributes (see Tables 2 and 3), the error in carcass classification reached 1.08 and 0.904 respectively. These results are very illustrative, since the errors reported in these tables, found by cross validation, are a good estimation of absolute differences for unseen cases. Thus, if our ML algorithms learn how to classify light carcasses and then try to grade conformation of standard carcasses, they will mistake not only the degree within a conformation class but also the class itself.

Additionally, our results strongly suggest that, as well as different attributes, classifiers apply different criteria when they classify different types of carcass. For instance, when the blockiness index was included in the training sets, this attribute was always taken into account in the rules used to assess conformation for each type of carcass. However, it seems as if the influence of this trait on classifiers would be different when dealing with a given carcass type. Classifiers apply SEUROP assumptions more rigorously when they classify standard carcasses. Our classifiers carry out light carcass assessments with a high degree of uncertainty. They cannot strictly follow SEUROP rules to obtain a conformation grading useful to market operators when they assess very young animals.

5. Implications

The aim of this paper was to test the validity of the official SEUROP bovine carcass classification system over the whole range of carcasses found in some important markets in the European Union in which the 'youthfulness' of the beef is highly regarded. To do so, we made use of the major advantages of employing AI techniques (Goyache, del Coz et al., 2001; Goyache, Bahamonde et

al., 2001): which a) do not need a linear relationship between the major attributes affecting the process that needs to be learned, and b) they can be used to ascertain the major attributes affecting process performance.

We found that the behavior of classifiers is quite different when they grade light or standard carcasses. Of course, carcass profiles and muscular development are important when taking decisions with respect to both types of carcasses (Goyache et al., in press), but the degree of uncertainty of classifiers when grading a light carcass is clearly higher than for standard carcasses. SEUROP assumptions are easier to apply to carcasses from animals that have reached a minimum degree of physiological maturity. When the assessed animal has not reached this degree of development, they need to take into consideration attributes characterizing body size or the degree of skeletal development. The SEUROP system may be criticized if applied to carcasses that do not come from 'young adult' animals. In the range of ages of animals producing light carcasses, we can find large differences in development and maturity. To obtain a 'SEUROP' classification for light carcasses, classifiers need to take into account attributes that are not included in the SEUROP definition. Consequently, we cannot expect the grades obtained for each type of carcass to be comparable. This could affect sire evaluation programs to improve carcass conformation scores that obtain data from markets presenting a great variety of ages and weights of slaughtered animals. A reconsideration of the SEUROP classification system for light carcasses is needed to clarify and standardize this specific beef market in the European Union.

Carcass classification has a significant importance in itself, as it is an important component of price negotiations between market operators. Since carcass shape and muscular development can vary substantially with the weight or age of the animals, methodologies to assess carcass conformation should take this into account. The European Mediterranean beef markets like light carcasses, and such carcasses are also significant in some northern European Union countries, such as The Netherlands, Germany or Belgium. The development of a specific methodology for grading light carcass conformation may be of interest for all the members of the European Union. The techniques we describe in this paper can be used to obtain objective information to adapt official conformation classification requirements to specific beef markets,so as to optimize market operations and consumer information.

Acknowledgements

References

Bahamonde, A., de la Cal, E. A., Ranilla, J., & Alonso, J. (1997). : Self-organizing symbolic learned rules. *Lecture Notes in Computer Science, (LNCS)*, *1240*, 536–545.

Belk, K. E., Scanga, J. A., Tatum, J. D., Wise, J. W., & Smith, C. G. (1998). Simulated instrument augmentation of USDA yield grade application to beef carcasses. *Journal of Animal Science*, *76*, 522.

Blum, A., & Langley, P. (1997). Selection of relevant features and examples in Machine Learning. *Artificial Intelligence Journal*, *1–2*, 245–271.

Cannell, R. C., Tatum, J. D., Belk, K. E., Wise, J. W., Clayton, R. P., & Smith, G. C. (1999). Dual component video image analysis system (VIASCAN™) as a predictor of beef carcass red meat yield percentage and for augmenting application of USDA yield grades. *Journal Animal Science*, *77*, 2942–2950.

Cubist (2000). Release 1.09. Available: http://www.rulequest.com/cubist-info.html.

De Boer, H., Dumont, B. L., Pomeroy, R. W., & Weniger, J. H. (1974). Manual on E.A.A.P. reference methods for the assessment of carcass characteristics in cattle. *Livestock Production Science*, *1*, 151–164.

Del Coz, J. J., Luaces, O., Quevedo, J. R., Alonso, J., Ranilla, J., & Bahamonde, A. (1999). Self-organizing cases to find paradigms. *Lecture Notes in Computer Sciences, (LNCS)*, *1606*, 527–536.

Díez, J., Goyache, F., Alonso, J., del Coz, J.J., Quevedo, J.R., López, S., Ranilla, J., Luaces, O., & Bahamonde, A. (2001). Utilización de técnicas de Inteligencia Artificial en la clasificación de canales bovinas. *Proceedings CAEPIA'01 IX Conferencia de la Asociación Española para la Inteligencia Artificial, Vol. II, 1229–1238*. Gijón, Spain, 14–16 November.

George, M. H., Dolezal, H. G., Tatum, J. B., Morgan, J. W., Wise, J. W., Calkins, C. R., Reagan, J. O., & Smith, C. G. (1996). *USDA yield grades, total body electrical conductivity and video image analysis technologies for predicting cutability of sides of steer/heifer carcasses. Beef Progam Report*. Colorado State University, CO, USA: Department of Animal Sciences.

Goyache, F., del Coz, J. J., Quevedo, J. R., López, S., Alonso, J., Ranilla, J., Luaces, O., Alvarez, I., & Bahamonde, A. (2001). Using artificial intelligence to design and implement a morphological assessment system in beef cattle. *Animal Science*, *73*, 49–60.

Goyache, F., Royo, L. J., Alvarez, I., & Gutiérrez, J. P. (2002). Testing a continuous variation in preweaning expression of muscular hypertrophy in beef cattle using field data. *Archives für Tierzucht*, *45*(2), 11–21.

Goyache, F., Bahamonde, A. Alonso, J., López, S., Alonso, J., del Coz J.J., Quevedo, J.R., Ranilla, J., Luaces, O., Alvarez, I., Royo, L. & Díez J. (2001). The usefulness of Artificial Intelligence techniques to assess subjective quality of products in the food industry. *Trend in Food Science and Technology 12*(10), 370–381.

Gutiérrez, J.P. & Goyache, F. (2002). Estimation of genetic parameters of type traits in Asturina de los Valles beef cattle breed. *Journal of Animal Breeding Genetics*, *119*(2), 93–100.

Hill, B. D., Jones, S. D. M., Robertson, W. M., & Major, I. T. (2000). Neural network modeling of carcass measurements to predict beef tenderness. *Canadian Journal of Animal Science*, *80*(2), 311–318.

Kempster, A. J. (1986). Estimation of the carcass composition of different cattle breeds and crosses from conformation assessments adjusted for fatness. *J. Agric. Sci. Camb*, *106*, 239–254.

Kempster, A. J., Cook, G. L., & Southgate, J. R. (1988). Evaluation of British, Canadian Holstein and beef breed×British Friesian steers slaughtered over a commercial range of fatness from 16-month and 24 month beef production system. 2. Carcass characteristics, and rate and efficiency of lean gain. *Animal Production*, *46*, 365–378.

Kira, K., & Rendell, L. A. (1992). A practical approach to feature selection. In *Proceedings of the 9th International Conference on Machine Learning*, (pp. 249–256). Aberdeen, Scotland: Morgan Kaufmann.

Luaces, O., del Coz, J. J., Quevedo, J. R., Alonso, J., Ranilla, J., & Bahamonde, A. (1999). Autonomous clustering for machine learning. *Lecture Notes in Computer Sciences, (LNCS)*, *1606*, 497–506.

More O'Ferral, G. J., & Keane, M. G. (1990). A comparison for live weight and carcass production of Charolais, Hereford and Friesian steer progeny from Friesian cows finished on two energy levels and serially slaughtered. *Animal Production*, *50*, 19–28.

Quevedo, J. R. & Bahamonde, A. (1999): Aprendizaje de Funciones Usando Inducción sobre Clasificaciones Discretas. *Proceedings CAEPIA'99 VIII Conferencia de la Asociación Española para la Inteligencia Artificial, Vol. I, 64–71*. Murcia, Spain, 16–19 November.

Quevedo J. R., Del Coz, J. J., Díez, J. (2001). Filtrando atributos para mejorar el aprendizaje automático. *Proceedings CAEPIA'01 IX Conferencia de la Asociación Española para la Inteligencia Artificial, Vol. I, 103–112*. Gijón, Spain, 14–16 November.

Smith G.C. (1999). New technologies for precision selection management and marketing of beef. Department of Animal Sciences, Colorado State University, CO, USA. Beef Program Report. Available: http://ansci.colostate.edu/ran/meat/newtechnologies.pdf.

Vallejo, M., Gutiérrez, J. P., Cima, M., Cañón, J., Alonso, L., Revuelta, J. R., & Goyache, F. (1992). Características de las canales de las razas bovinas asturianas. III.—Valoración cuantitativa y predicción de la composición tisular de canales en la raza Asturiana de los Valles. *Archivos de Zootecnia*, *42*, 29–90.

Wettschereck, D., Aha, D. W., & Mohri, T. (1997). A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*, *11*, 273–314.