



**This document is a postprint version of an article published in Meat Science © Elsevier after peer review. To access the final edited and published work see <https://doi.org/10.1016/j.meatsci.2018.04.011>**

# 1 On-line Ham Grading using pattern recognition models based 2 on available data in commercial pig slaughterhouses

3 Gerard Masferrer<sup>ab</sup>, Ricard Carreras<sup>b</sup>, Maria Font-i-Furnols<sup>c</sup>, Marina Gispert<sup>c</sup>, Pere Marti-Puig<sup>a</sup>, Moises Serra<sup>a</sup>

4 <sup>a</sup>Information and Digital Technology Department, UVic-UCC, Vic (Barcelona)

5 <sup>b</sup>Mafrika.SA, Paratge Can Canals Nou, S/N 08250 Sant Joan de Vilatorrada (BCN)

6 <sup>c</sup>IRTA-Product Quality, 17121 Monells, Girona, Spain

---

## 8 Abstract

9 The thickness of the subcutaneous fat in hams is one of the most important factors for the dry-  
10 curing process and largely determines its final quality. This parameter is usually measured in  
11 slaughterhouses by a manual metrical measure to classify hams. The aim of the present study was  
12 to propose an automatic classification method based on data obtained from a carcass automatic  
13 classification equipment (AutoFom) and intrinsic data of the pigs (sex, breed, and weight) to  
14 simulate the manual classification system. The evaluated classification algorithms were decision  
15 tree, support vector machines (SVM), k-nearest neighbour and discriminant analysis. A total of  
16 4000 hams selected by breed and sex were classified as thin (0-10mm), standard (11-15 mm),  
17 semi-fat (16-20 mm) and fat (>20 mm). The most reliable model, with a percentage of success of  
18 73%, was SVM with Gaussian kernel, including all data available. These results suggest that the  
19 proposed classification method can be a useful online tool in slaughterhouses to classify hams.

20 **Keywords** dry-cured hams; ham-fat grading; subcutaneous fat thickness; pattern recognition

---

## 22 1. Introduction

23 Ham is one of the most valued product in pork meat industry. This primal cut represents between  
24 25 and 30 percent of the carcass (Cisneros, Ellis, & McKeith, 1996; Gispert et al., 2007) and is  
25 the basis of different regional specialities focused on preserving and flavouring raw meat  
26 (Dirinck, Van Opstaele, & Vandendriessche, 1997). Those specialities include different  
27 techniques such as salting dry-cured ham, smoking or wet curing. Some examples are  
28 Westphalian ham in Germany, Prosciutto in Italy, and Jamon Serrano in Spain.

29 The Subcutaneous Fat Thickness (SFT) in hams determines, among other factors, which is the  
30 best process for the ham to be submitted. Hams with low subcutaneous fat have a high lean meat  
31 percentage (LMP) and are more appropriate to be processed as raw or cooked meat while hams  
32 with higher subcutaneous fat are more appropriate to be cured or smoked.

33 Moreover, the SFT determines the optimum curing time (Bosi, Russo, & Paolo, 2004), which is  
34 directly related to the quality of the final product (Čandek-Potokar & Škrlep, 2012). Therefore,  
35 classify the ham according to the SFT is crucial to get the maximum benefit of the product, in  
36 economic and quality terms.

37 The thickness of the subcutaneous fat is determined by several factors, among which can be  
38 highlighted the breed (Gispert et al., 2007; Wood et al., 2004), the sex (Font-i-Furnols et al., 2012;  
39 Gispert et al., 2010), the slaughter weight (Fàbrega et al., 2011; Latorre, García-Belenguier, &  
40 Ariño, 2008) and the diet (Realini et al., 2010; Tous et al., 2014; Wood et al., 2004). Regarding  
41 the breed, there are leaner breeds, as would be the Pietrain and other fatter breeds such as the  
42 Duroc (Cilla et al., 2006; Edwards, Bates, & Osburn, 2003). In terms of sex, females tend to  
43 deposit more subcutaneous fat than males (Gispert et al., 2010; Wood, Enser, Whittington,  
44 Moncrieff, & Kempster, 1989). Moreover, the castration, especially surgical but also  
45 immunological, also contributes to deposit more subcutaneous-fat compared with entire male pigs  
46 (Gispert et al., 2010; Wood et al., 2008).

47 Nowadays slaughterhouses have different methods to estimate the SFT of hams. One of the most  
48 used method is the visual system based on a metrical measure of the SFT over the *Gluteus medius*  
49 muscle, similar to ZP (Zwei-Punkte Messverfahren) measures, used to determine carcass LMP  
50 (Daumas, 2011; Font-i-Furnols et al., 2016). Indeed, the carcass LMP is a parameter widely used  
51 in slaughterhouses as the current EU legislation establishes it as compulsory for carcasses  
52 classification. There are different methods to determine LMP based, predominantly, on the  
53 existing relationship of thickness between fat and muscle in several parts of the carcass (Font i  
54 Furnols & Gispert, 2009).

55 Obtaining these measures manually is unsuitable in slaughter plants with medium/high speed line,  
56 therefore the most used methods to determine LMP are semiautomatic systems based on  
57 reflectance penetration probes, as for instance the Fat-O-Meat'er (FOM; Frontmatec Smørum  
58 A/S, Herlev, Denmark) or the Hennessy Grading Probe (HGP; Hennessy Grading System Ltd.,  
59 Auckland, New Zealand), which determine fat and muscle thickness at a defined anatomical  
60 position and use them to estimate carcass LMP. Alternatively, there are non-invasive and fully  
61 automatic systems such as AutoFom (Frontmatec Smørum A/S, Herlev, Denmark) which is based  
62 on three-dimensional ultrasonic systems, or VCS 2000 (e + V Technology GmbH, Oranienburg,  
63 Germany) that extracts LMP by processing and analysing images (Font i Furnols & Gispert,  
64 2009). Some of these devices also can estimate several SFT at the loin and at the ham level. For  
65 instance, AutoFom, provides several SFT parameters of the ham like *fatham2* (minimum  
66 subcutaneous fat plus skin thickness measured with a ruler over the muscle *Gluteus medius*) and  
67 *fatham3* (thickness of the subcutaneous fat plus skin measured with a ruler, perpendicularly to the

68 skin, at the cranial part of muscle *Gluteus medius*).

69 Other systems, such as thermography technology have been proposed to classify the hams  
70 according to the SFT, being the hams with lower fat cover the ones that display a significantly  
71 warmer average temperature surface (Nanni Costa et al., 2010). Also computed tomography has  
72 been used in experimental conditions to determine the fat thickness at different anatomical  
73 positions mainly in the loin region (Lucas et al., 2017) although it could also been used in the  
74 ham region as has been done in live pigs (Carabús et al., 2014).

75 Nowadays a certain amount of data is collected in the slaughter line like gender and carcass  
76 weight, but also much other information from the productive chain is available such as breed,  
77 diet, transport and farm conditions, medication and castration (if done). In this context, with all  
78 this available data it is possible to take technical and commercial real-time decisions to better  
79 classify products and maximise profits. Therefore, our hypothesis is that complementing the  
80 Autofom-III set of estimated parameters with those additional ones could be used to improve the  
81 ham classification rate according to the SFT.

82 To carry out this classification it is possible to use classifiers. A classifier is an algorithm used to  
83 assign an unlabelled incoming element in a known category based on certain characteristic  
84 information of that element. These algorithms need to perform a learning stage. There are two  
85 types of primary learning strategies: supervised learning which elaborates a mathematical  
86 function (hypothesis) from previously labelled training data and unsupervised learning which  
87 does not have a training package that allows knowing the data labels, so it is necessary to use  
88 grouping techniques that try to build these labels. Among supervised algorithms, some of the most  
89 widespread are *Decision Trees*, *K-Nearest Neighbours (KNN)*, *Linear and Nonlinear*  
90 *Discriminant Analysis (LDA/nLDA)* and *Support Vector Machine (SVM)* (Bishop, 2006).  
91 Between the unsupervised classifiers the most popular strategies are the clustering which includes  
92 the Hierarchical and k-Means clustering algorithms.

93 Thus, the objectives of this study are: (1) To apply and assess different supervised classification  
94 techniques (Decision trees, kNN, SVN, LDA/nLDA) to predict the classification of hams  
95 according to SFT by combining data form Autofom III and intrinsic data from the animal, (2) to  
96 evaluate the impact of each predictor in the accuracy of ham classification, and (3) to evaluate  
97 several combinations of predictors available in different slaughterhouses scenarios and to  
98 compare them.

99

## 100 2. Material and Methods

101 **2.1 Animals and facilities. The dataset construction**

102 This study was carried out with data obtained during May 2016 from pigs fattened in Spanish  
 103 commercial farms and slaughtered in a commercial slaughterhouse (MAFRICA S.A.) located in  
 104 Sant Joan de Vilatorrada, Catalonia, Spain. All farms were less than 200 km far from the  
 105 slaughterhouse and pigs were transported using trucks in groups (usually of between 80 and 220  
 106 animals). Once in the slaughterhouse pigs rested into lairage pens between 2 and 4 hours before  
 107 being slaughtered.

108 This slaughterhouse works five days per week slaughtering a mean of 1700 pigs per day, obtaining  
 109 more than 32000 carcasses per month. **A total of 4000 carcasses were selected for this study**  
 110 **according to their breed and sex in order to ensure a representative sample regarding fat thickness.**  
 111 Those carcasses were selected according to their sex: 60.6% females, 19.4% entire males and  
 112 20.0% castrated males and according to their genetics: 51.9% (Large White × Landrace) ×  
 113 Piétrain, 38.3% were (Large White x Landrace) x Duroc and 9.8% (Large White x Landrace) x  
 114 (Duroc x Landrace). **Table 1 shows the mean weight of the cold carcass and the fat thickness of**  
 115 **the pigs according to the breed and sex. Fat thickness parameter is given by the ultrasound**  
 116 **AutoFom-III system and corresponds to the parameter F34 that is described as the fat thickness**  
 117 **at 60 mm in the mid-line between the 3rd and the 4th last rib.**

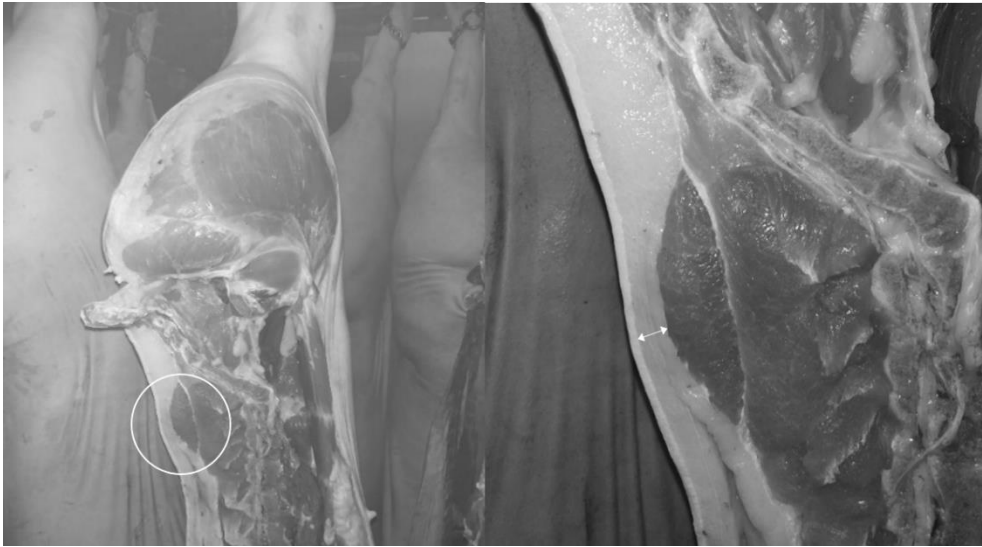
118 Table 1. The cold carcass weight (mean ± s.d; kg) and the fat thickness at 60 mm in the mid-line between the 3<sup>rd</sup> and  
 119 the 4<sup>th</sup> last rib (mean ± s.d.; mm) of 4000 carcass according to breed and sex.

<b>BREED</b>	<b>n</b>	<b>WEIGHT (mean ± s.d; kg)</b>	<b>FAT THICKNESS (mean ± s.d.; mm)</b>
(Large White × Landrace) × Piétrain	2077	81.80 ± 8.16	15.39 ± 4.10
(Large White x Landrace) x Duroc	1531	93.76 ± 10.69	24.55 ± 5.56
(Large White x Landrace) x (Duroc x Landrace)	392	85.92 ± 9.02	18.60 ± 5.20
<b>SEX</b>			
Female	2289	85.49 ± 10.00	17.59 ± 5.63
Castrated	1315	90.97 ± 11.54	23.51 ± 6.19
Entire male	396	80.38 ± 7.91	14.29 ± 3.22

120

121 Pigs were slaughtered after stunning with CO<sub>2</sub> (90%) for 2 min. After scalding they were totally  
 122 **monitored** using the ultrasound AutoFom-III system. Then pigs were eviscerated and splitted  
 123 according to standard commercial procedures using an automatic robotic system. **After that, the**  
 124 **two half-carcasses were weighted and an experimented operator visually determined the sex of**  
 125 **the pig (female, entire male or castrated male) and classified the left half carcass according to**  
 126 **minimal fat depth over muscle *gluteus medius* which is shown in Fig. 1. Classes were established**  
 127 **based on the measures shown in Table 2. The operator had a pattern, based on these classes, that**

128 was used to visually compare and determine in which of the four ham classes (HC) each ham was  
 129 classified.



130

131 *Fig. 1. Representation of the section used and the measure performed by an expert operator to measure the minimal*  
 132 *fat thickness over muscle gluteus medius to obtain the classification target.*

133 *Table 2. Carcass classification according to minimal fat thickness over muscle gluteus medius based on a metrical*  
 134 *measure with a ruler*

Ham_Class (HC)	Fat depth (mm)
(1)- Thin	<10
(2)- Standard	Between < 10 and 15
(3)- Semi-fat	Between < 15 and 20
(4)- Fat	> 20

135

## 136 2.2 Dataset predictors

137 AutoFom-III predicts carcass LMP and seven other variables (Table 3) from 48 parameters  
 138 obtained from the scanning. Nevertheless, a more accurate handmade classification process of the  
 139 ham is required for commercial purposes. With the aim of improving classification rates the eight  
 140 estimations provided by AutoFom-III, that are going to be used as predictors, are complemented  
 141 with three more predictors obtained in the production line (sex, breed, and weight) (Table 3). The  
 142 extended set of 11 predictors was used as the input of automatic classification systems applying  
 143 pattern recognition techniques to assess different classifiers.

144 *Table 3. The eleven predictors used as the input of automatic classification systems*

Predictor	Description
<b>Autofom III</b>	
<b>LMP</b>	Lean Meat Percentage

<b>F34</b>	According to the official formula, the subcutaneous fat thickness at 60 mm in the mid-line between the 3 <sup>rd</sup> and the 4 <sup>th</sup> last rib. (mm)
<b>M34</b>	According to the official formula, muscle thickness at 60 mm in the mid-line from the 3 <sup>rd</sup> to the 4 <sup>th</sup> last rib. (mm)
<b>F_GM1</b>	The minimum subcutaneous fat plus skin thickness measured with a ruler over the muscle <i>Gluteus medius</i> (mm)
<b>F_GM2</b>	The thickness of the subcutaneous fat plus skin measured with a ruler, perpendicularly to the skin, at the cranial part of muscle <i>Gluteus medius</i> . (mm)
<b>WGT_H</b>	Total weight of the ham (kg)
<b>WGT_HWB</b>	Ham's weight without bone (kg)
<b>WGT_HLM</b>	Total weight of the lean meat of the ham (kg)

**Production line**

<b>SEX</b>	Sex of animals ( <b>females, entire males and castrated males</b> )
<b>BREED</b>	Crossbreed (( <b>Large White x Landrace</b> ) x <b>Pietrain</b> , , ( <b>Large White x Landrace</b> ) x <b>Duroc</b> , and ( <b>Large White x Landrace</b> ) x ( <b>Duroc x Landrace</b> ))
<b>WGT</b>	Cold carcass weight ( <b>kg</b> )

145

146 Finally, the HC parameter (1, 2, 3 or 4; see Table 2) used as a response was scored by an expert  
 147 operator and is referred to the manual metrical measure to classify hams according to the thickness  
 148 of the fat at the point shown in Fig.1.

149

150 **2.3 Predictors and classifiers evaluated**

151 A preliminary study was performed to evaluate the potential of each predictor individually to  
 152 forecast the HC classification. Therefore, each single predictor was only considered to feed each  
 153 of the classifiers to obtain the response. All classifiers were evaluated in terms of the accuracy  
 154 which is defined as the number of correct predictions divided by the number of total predictions.

155 Moreover, the impact in the prediction of HC when taking different combinations of predictors  
 156 as inputs in the classifier was also assessed in terms of the accuracy. The aim of this assessment  
 157 was to compare the predictability of the classifiers when trained with only the single input LMP,  
 158 and when other predictors are incorporated, such as the combinations of LMP and SEX or LMP  
 159 and BREED (see Table 4) for all the combinations. These combinations were chosen according  
 160 to the different slaughterhouse scenarios described below.

161 *Table 4. Predictors included in each dataset*

**Predictors used as inputs**

<b>Datasets</b>	<b>LMP<sup>1</sup></b>	<b>SEX<sup>2</sup></b>	<b>WGT<sup>3</sup></b>	<b>BREED<sup>4</sup></b>	<b>F34<sup>5</sup></b>	<b>M34<sup>6</sup></b>	<b>F_GM1<sup>7</sup></b>	<b>F_GM2<sup>8</sup></b>	<b>WGT_H<sup>9</sup></b>	<b>WGT_HWB<sup>10</sup></b>	<b>WGT_HLM<sup>11</sup></b>
<b>D1</b>	X										
<b>D2</b>	X	X									
<b>D3</b>	X		X								
<b>D4</b>	X			X							
<b>D5</b>	X	X	X	X							
<b>D6</b>	X	X	X	X	X	X					

<b>D7</b>	X	X	X	X	X	X	X	X	X	X	X
-----------	---	---	---	---	---	---	---	---	---	---	---

163 <sup>1</sup>LMP (Lean Meat Percentage); <sup>2</sup>SEX (females, entire males and castrated males); <sup>3</sup>WGT (warm carcass weight); <sup>4</sup>BREED ( (Large  
164 White x Landrace) x Pietrain, (Large White x Landrace) x Duroc and (Large White x Landrace) x (Duroc x Landrace)); <sup>5</sup>F34  
165 (subcutaneous fat thickness at 60 mm in the mid-line between the 3<sup>rd</sup> and the 4<sup>th</sup> last rib); <sup>6</sup>M34 (loin depth in mm measured at 60 mm  
166 from the midline between the 3<sup>rd</sup> and the 4<sup>th</sup> last rib); <sup>7</sup>F\_GM1 (minimum subcutaneous fat plus skin thickness measured with a ruler  
167 over the muscle *Gluteus medius*); <sup>8</sup>F\_GM2 (thickness of the subcutaneous fat plus skin measured with a ruler, perpendicularly to the  
168 skin, at the cranial part of muscle *Gluteus medius*); <sup>9</sup>WGT\_H (total weight of the ham); <sup>10</sup>WGT\_HWB (ham's weight without bone);  
169 <sup>11</sup>WGT\_HLM (total weight of the lean meat of the ham).

170 According to the Commission Delegated Regulation (EU) 2017/1182, it is mandatory in all the  
171 slaughterhouses to classify pig carcasses by means of its LMP. Therefore the combination **D1**  
172 (Table 4) is available in the production line of all slaughterhouses.

173 As more procedures are added in the slaughtering line, more predictors could be obtained in real-  
174 time such as SEX, BREED, and WGT. Those additional predictors can be incorporated as inputs  
175 in the classifiers, as it has been done from **D2** to **D5**.

176 Combination **D6** considers the addition of predictors F34 and M34 that are provided by AutoFOM  
177 III. These predictors have been chosen because they can be assessed using other classification  
178 systems like Fat-O-Meat'er- FOM (Kempster, Chadwick, & Jones, 1985). Finally, **D7** takes all  
179 additional information given by AutoFOM III (predictors **F\_GM1**, **F\_GM2**, **WGT\_H**,  
180 **WGT\_HWB**, and **WGT\_HTL**) (Table 4.).

## 181 **2.4 Statistical analysis**

182 To train each classifier four sets of 1000 samples of each HC class were randomly selected from  
183 the total of **31188** ones to form a balanced group of 4000 samples. Afterwards, to prevent the  
184 classifier overfitting, a *5-Fold cross-validation* method was used (Bishop, 2006) dividing the  
185 dataset into 5 subsets, and for 5 times one of the 5 subsets was used as test set and the other 4  
186 subsets get together to form a training set and the average error across all 5 trials was computed.

187 All classifiers were evaluated in terms of the accuracy (number of correct predictions divided by  
188 the number of total predictions).

189 A set of well-known classifier techniques was evaluated (Bishop, 2006): (1) Decision Trees: this  
190 type of algorithm is based on the construction of an automatic diagram of branches that appear  
191 according to the available data and the specific weight of each parameter. This algorithm was  
192 used with 4, 20 and 100 maximum split-levels; (2) Support Vector Machines (SVM): a  
193 discriminative classifier that separates classes by a hyperplane. The SVM algorithm is based on  
194 finding the optimal separating hyperplane that gives the largest minimum distance between the  
195 classes of the training data. This algorithm was used with four different kernels - linear, quadratic,



196 cubic and Gaussian (Burgess, 1998; Vapnik & Chervonenkis, 1964); (3) K-Nearest Neighbour  
197 Classifiers (*K*-NN): a non-parametric supervised classifier based on the comparison of a sample  
198 against the *K* samples which most resemble assigning the most abundant class (Cover & Hart,  
199 1967). This algorithm was used with six different configurations; (4) Discriminant Analysis with  
200 linear (Balakrishnama & Ganapathiraju, 1998; Fisher, 1936) and quadratic configurations based  
201 on finding a linear or quadratic combination of parameters that characterise or separates two or  
202 more classes.

203 MATLAB and Signal Processing Toolbox™ (Matlab R2016b; The MathWorks, Inc, 1988–2016)  
204 have been used to develop and test all the models and algorithms.

### 205 **3. Results and discussion**

206 Table 5 shows the accuracy of 17 classification models when a single predictor is taken as input.  
207 These classification models allow interpreting the results as a measure of the impact that each  
208 predictor by itself has in the forecast. Accuracy oscillates between 15 and 68% depending on the  
209 predictor and the type of classifier. Predictors **F\_GM1** and **F\_GM2** obtain the best results of  
210 accuracy in most of the classifiers, outperforming the results obtained by LMP. F34 also achieves  
211 good results regarding accuracy, however, in this case, the results are more dependent on the  
212 classifier type. Those results were foreseeable as predictors **F\_GM1**, **F\_GM2** and **F34** provide  
213 information about a direct measure of fat thickness in two points of the ham and in one point of  
214 the loin, respectively. Indeed, they are physically related to the handmade measure taken by an  
215 expert operator who assigns the HC class. On the other hand, predictors such as SEX, WGT and  
216 BREED can be good predictors to classify the hams correctly but largely depends on the type of  
217 classifier.

218 The highest and the lowest accuracy values for each predictors' dataset are presented **in bold and**  
219 **underlined**, respectively. The best results of predictors **F\_GM1**, **F34**, **F\_GM2** and LMP predicted  
220 the HC class with an accuracy between 63 and 68%. Moreover, predictors BREED, WGT and  
221 SEX predicted the HC class with an accuracy between 42 and 48%. Finally, the rest of predictors,  
222 had an accuracy below 37%.

223 In general, SVM Medium Gaussian or Coarse Gaussian or the Fine worked better when predictors  
224 are lean or fat parameters while SVM Cubic is one of the worst. This result persists in all  
225 predictors used but the interpretation about the relation of SVM kernels and the dataset is not  
226 clear.

227 When weight predictors are used, linear or quadratic discriminant analysis, and also Medium  
228 Gaussian, Coarse Gaussian and fine SVM produce the highest accuracy. These results suggest  
229 that continuous variables, such as the weight, improve the accuracy of more complex algorithms

230 while categorical variables fits better with more simple algorithms. Sex and breed have higher  
 231 accuracy when decision trees and SVM are used and discriminant analysis for breed. We can  
 232 hypothesize than sex and breed obtain higher accuracy in decisions trees because, in the dataset,  
 233 they are only three breed classes (Table 1). According to the results of (Gispert et al 2007), there  
 234 is a clear relation between breed and SFT that could be easily formalized in simple decision trees.  
 235 Similar relations have been found for sex (Font-i-Furnols et al., 2012; Gispert et al., 2010). The  
 236 lowest accuracy is for the kNN approach. We can observe that for the classification of ham is  
 237 usually more relevant breed than weight, and in turn, weight than sex.

238 *Table 5 The Accuracy (in percentage) to predict the Ham Classification (HC) based on the thickness of the*  
 239 *subcutaneous fat of the ham for each classifier when a single predictor is considered.*

Classifiers	Predictors										
	LMP <sup>1</sup>	F34 <sup>5</sup>	M34 <sup>6</sup>	F_GMI <sup>7</sup>	F_GM2 <sup>8</sup>	WGT_H <sup>9</sup>	WGT_HWB <sup>10</sup>	WGT_HTL <sup>11</sup>	SEX <sup>2</sup>	WGT <sup>3</sup>	BREED <sup>4</sup>
<b>Decision Trees</b>											
Simple tree	62	65	36	68	63	33	33	32	42	44	48
Medium tree	61	65	36	67	64	33	32	32	42	43	48
Complex tree	61	63	36	65	62	32	32	29	42	42	48
<b>Support Vector Machines</b>											
Linear	52	58	27	59	51	28	27	26	40	35	47
Quadratic	31	38	25	40	44	25	25	26	42	27	48
Cubic	15	19	25	35	22	24	23	24	42	19	48
Fine	63	65	36	68	65	34	33	32	42	44	48
Medium Gaussian	63	65	36	68	65	35	33	32	42	44	48
Coarse Gaussian	63	65	36	68	64	34	33	32	42	44	48
<b>K-Nearest Neighbours</b>											
Fine	36	53	30	57	53	27	28	25	25	26	25
Medium	58	62	33	65	63	31	32	28	25	33	25
Coarse	62	65	35	68	64	32	32	31	27	41	25
Cosine	25	25	25	25	25	25	25	25	25	25	25
Cubic	58	61	33	65	63	32	32	28	25	32	25
Weighted	57	56	31	60	55	29	30	27	25	32	25
<b>Discriminant analysis</b>											
Linear	62	65	36	68	64	34	33	32	34	44	48
Quadratic	61	64	36	68	63	35	34	32	39	44	48

241 <sup>1</sup>LMP (Lean Meat Percentage); <sup>2</sup>SEX (Females, entire males and castrated males); <sup>3</sup>WGT (warm carcass weight); <sup>4</sup>BREED ( (Large  
242 White x Landrace) x Pietrain, (Large White x Landrace) x Duroc and (Large White x Landrace) x (Duroc x Landrace)); <sup>5</sup>F34  
243 (subcutaneous fat thickness at 60 mm in the mid-line between the 3<sup>rd</sup> and the 4<sup>th</sup> last rib); <sup>6</sup>M34 (loin depth in mm measured at 60 mm  
244 from the midline between the 3<sup>rd</sup> and the 4<sup>th</sup> last rib); <sup>7</sup>F\_GM1 (minimum subcutaneous fat plus skin thickness measured with a ruler  
245 over the muscle *Gluteus medius*); <sup>8</sup>F\_GM2 (thickness of the subcutaneous fat plus skin measured with a ruler, perpendicularly to the  
246 skin, at the cranial part of muscle *Gluteus medius*); <sup>9</sup>WGT\_H (total weight of the ham); <sup>10</sup>WGT\_HWB (ham's weight without bone);  
247 <sup>11</sup>WGT\_HLM(total weight of the lean meat of the ham). In bold he highest value for each dataset; Underlined lowest value for each  
248 dataset.

249 Table 6 shows the accuracy of each classifier according to the data set configurations that are  
250 more commonly available in different slaughterhouse scenarios, as described in section 2.3, Table  
251 4. As commented in section 2.4 classifiers were obtained and validated with cross validation with  
252 the 4000 carcasses. In addition, although the 27188 were a non-balanced data set in terms of HC,  
253 (i.e. 16920 (thin), 6074 (standard), 4003 (semi-fat) and 191 (fat)) the classifiers were also  
254 validated using this dataset and accuracy of the results was similar to the obtained by cross  
255 validation (data not shown).

256 *Table 6 Accuracy (in percentage) of each classification model with different dataset configurations<sup>1</sup> used to train*  
257 *models.*

Classifiers	Datasets						
	D1	D2	D3	D4	D5	D6	D7
<b>Decision Trees</b>							
Simple tree	62	62	63	62	64	65	68
Medium tree	61	64	65	62	67	68	70
Complex tree	61	63	64	61	68	67	68
<b>Support Vector Machines</b>							
Linear	52	61	<b>67</b>	<b>63</b>	<b>69</b>	<b>71</b>	71
Quadratic	31	45	60	49	68	<b>71</b>	72
Cubic	<u>15</u>	<u>32</u>	<u>33</u>	<u>37</u>	68	<b>71</b>	69
Fine	<b>63</b>	<b>65</b>	66	<b>63</b>	68	69	69
Medium Gaussian	<b>63</b>	<b>65</b>	<b>67</b>	<b>63</b>	<b>69</b>	70	<b>73</b>
Coarse Gaussian	<b>63</b>	64	<b>67</b>	<b>63</b>	68	<b>71</b>	71
<b>K-Nearest Neighbours</b>							
Fine	36	43	56	42	59	<u>61</u>	<u>62</u>
Medium	58	61	64	58	65	66	68
Coarse	62	64	56	62	65	68	67
Cosine	25	61	56	58	65	68	68
Cubic	58	61	65	57	65	68	68

<b>Weighted</b>	57	59	60	<b>63</b>	63	67	68
<b>Discriminant Analysis</b>							
<b>Linear</b>	62	56	<b>67</b>	59	64	67	70
<b>Quadratic</b>	62	58	65	54	<u>55</u>	63	66

258 <sup>1</sup>In bold the highest value for each dataset; Underlined the lowest value for each dataset.

259 <sup>1</sup> See Table 2 for description of the inputs included as predictors in each dataset studied (from D1 to D7).

260 The first column shows the results obtained using LMP as a single predictor. The highest value  
261 (stood out in bold, Table 6) of the different classifiers for dataset configurations. D2 and D3 show  
262 a positive impact on most of the classifiers accuracy due to the incorporation of SEX and WGT  
263 predictors, respectively, compared with D1. Moreover, dataset configuration D4, in which  
264 BREED predictor has been incorporated, the accuracy improves just in some of the classifiers,  
265 such as SVM Linear and KNN Cosine. Predictor WGT seems to better complement LMP than  
266 SEX and BREED according to results obtained by Latorre, García-Belenguer, & Ariño (2008).

267 As a general rule, SVM Coarse, SVM Medium Gaussian and SVM Fine obtain the highest  
268 accuracy when only one or two predictors are used (D1 to D4) compared with the other classifiers.  
269 Moreover, when more predictors are used, all the SVM classifiers produce better results than the  
270 other classifier techniques. In addition, the more predictors are added, the better results are  
271 obtained with the most sophisticated classifiers, such as SVMs with complex kernels.

272 When SEX, WGT and BREED predictors complement LMP (D5) the accuracy of SVM Medium  
273 Gaussian, one of the classifiers with the highest accuracy in D1, increases a 6%, obtaining an  
274 accuracy value of 69%. Furthermore, the SVM Linear with D5, also obtain an accuracy value of  
275 69% increasing by 17% with respect to D1.

276 D6 dataset configuration incorporates to D5 predictors F34 and M34 obtained by Autofom.  
277 Configuration D7 has all available predictors (see section 2.2), obtained through the use of  
278 Autofom and intrinsic characteristics of the animal. In configurations D6 and D7, the classifiers  
279 obtain a percentage of accuracy between 61 and 73%. As expected, D7 configuration obtains the  
280 best performance. Regarding the classifiers, the SVM Medium Gaussian reached the best result  
281 with a percentage of accuracy of 73%.

282 When comparing models obtained from datasets D6 and D7, in average, there is a 1.0% of  
283 prediction improvement. It is suggested that the improvement is not greater because the added  
284 parameters are closely correlated with the previous ones. For instance, the five new predictors  
285 (F\_GM1, FGM2, WGT\_H, WGT\_HWB, WGT\_HLM) introduced in the models with input  
286 dataset D7 are highly correlated with predictors WGT and/or F34, present in dataset D6. However,  
287 although an increase of 1.0% does not represent a great improvement in terms of percentage of

288 success, it can mean a significantly improvement in the benefits of a company. Misclassifications  
 289 of a ham in a lower category, in terms of subcutaneous fat, could incur in losses of more than 30%  
 290 in the final sale price.

True class	1	79.0%	19.9%	0.9%	0.2%
	2	19.7%	65.6%	14.1%	0.6%
	3	1.2%	13.9%	65.8%	19.1%
	4	0.1%	0.6%	19.2%	80.1%
		1	2	3	4
		Predicted class			

291

292 *Fig. 2 Confusion Matrix of the best accuracy models obtained using SVM medium Gaussian model trained with all*  
 293 *data available (D7). The results are given in percentage.*

294 Fig.2 shows the confusion matrix obtained by SVM Medium Gaussian model developed using  
 295 **D7**. The accuracy of HC classes 1 (79.0%) and 4 (80.1%) are higher than the accuracy of HC  
 296 classes 2 (65.6%) and 3 (65.8%). When classes based on a metric threshold are used, extreme  
 297 classes tend to be better classified.

298 The percentage of samples that are incorrectly classified into one of the adjacent categories varies  
 299 between 13.9%-19.9% (Fig. 2.). It should be noted that some of these samples fall very close to  
 300 the decision thresholds and, in those cases, the classification is particularly difficult.

301 Moreover, only less than 3.6% of the samples are misclassified in not adjacent categories. Indeed,  
 302 it can be concluded that 96.7% of the 27.4% of misclassified samples correspond to samples  
 303 classified into adjacent categories.

304 As explained before, all the models are developed in order to predict the classification of the hams  
 305 by an expert operator. Indeed, in this study the human classification methodology is used as  
 306 “golden standard” despite the fact that this methodology presents some difficulties such as  
 307 operator fatigue (Font-i-Furnols et al., 2016; Olsen et al., 2007) but also the evaluation of the fat  
 308 thickness after the carcass being split down by an industrial robot (the carcasses are not precisely  
 309 split down in the same way). Therefore, misclassifications of the models do not always mean that  
 310 the model is classifying wrong, they are just explaining that the model classification does not  
 311 match with the human classification.

312 Nowadays, the SVM Medium Gaussian model is applied in MAFRICA S.A. slaughterhouse. It  
 313 is observed an accuracy improvement which is not currently quantified. Our working hypothesis

314 is that automatic classification improves manual classification because decision making is  
315 objective and operator fatigue are eliminated.

#### 316 **4. Conclusions**

317 Pattern recognition models, based on data usually available on slaughterhouses, can be used to  
318 classify the hams according to the thickness of the subcutaneous fat, and this classification can  
319 emulate the manual system with an effectivity of 73%. This result suggests that pattern  
320 recognition models can be a useful online tool to increase slaughterhouses' benefits because more  
321 accurate classification increases optimization of the ham processing.

#### 322 **Acknowledgements**

323 “Financial support from the Agency for Management of University and Research Grants  
324 (AGAUR) of the Catalan Government to Gerard Masferrer is grateful”. CERCA Programme  
325 (Generalitat de Catalunya) is also acknowledged.

#### 326 **References**

- 327 Andersen, H. (n.d.). AutoFom III TM -using data for genetic development and payment.  
328 Retrieved from [http://www.carometec.com/images/zoo/pdf/brochures/Application-](http://www.carometec.com/images/zoo/pdf/brochures/Application-Note_English.pdf)  
329 [Note\\_English.pdf](http://www.carometec.com/images/zoo/pdf/brochures/Application-Note_English.pdf)
- 330 Balakrishnama, S., & Ganapathiraju, A. (1998). Linear Discriminant Analysis - a Brief Tutorial.  
331 *Compute, 11*, 1–9.  
332 <http://doi.org/http://www.isip.piconepress.com/publications/reports/1998/isip/lda/>
- 333 Bishop, C. M. (2006). *Pattern Recognition and Machine Learning. Pattern Recognition* (Vol.  
334 4). <http://doi.org/10.1117/1.2819119>
- 335 Bosi, P., Russo, V., & Paolo, B. (2004). The production of the heavy pig for high quality  
336 processed products. *J.ANIM.SCI*, 3(21), 309–32145. <http://doi.org/10.4081/ijas.2004.309>
- 337 Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data*  
338 *Mining and Knowledge Discovery*, 2(2), 121–167.  
339 <http://doi.org/10.1023/A:1009715923555>
- 340 Candek-Potokar, M., & Skrlep, M. (2012). Factors in pig production that impact the quality of  
341 dry-cured ham: a review. *Animal : An International Journal of Animal Bioscience*, 6(2),  
342 327–38. <http://doi.org/10.1017/S1751731111001625>
- 343 Carabús, A., Gispert, M., Brun, A., Rodríguez, P., & Font-i-Furnols, M. (2014). In vivo  
344 computed tomography evaluation of the composition of the carcass and main cuts of

345 growing pigs of three commercial crossbreeds. *Livestock Science*, 170, 181–192.  
346 <http://doi.org/10.1016/j.livsci.2014.10.005>

347 Cilla, I., Altarriba, J., Guerrero, L., Gispert, M., Martínez, L., Moreno, C., ... Roncalés, P.  
348 (2006). Effect of different Duroc line sires on carcass composition, meat quality and dry-  
349 cured ham acceptability. *Meat Science*, 72(2), 252–260.  
350 <http://doi.org/10.1016/j.meatsci.2005.07.010>

351 Cisneros, F., Ellis, M., & McKeith, F. (1996). Influence of slaughter weight on growth and  
352 carcass characteristics, commercial cutting and curing yields, and meat quality of barrows  
353 and gilts from two genotypes. *Journal of Animal*. Retrieved from  
354 <https://dl.sciencesocieties.org/publications/jas/abstracts/74/5/925>

355 Commission Delegated Regulation (EU) 2017/1182 of 20 April 2017 supplementing Regulation  
356 (EU) No 1308/2013 of the European Parliament and of the Council as regards the Union  
357 scales for the classification of beef, pig and sheep carcasses and as regards the reporting of  
358 market prices of certain categories of carcasses and live animals. Official Journal of the  
359 European Union 4.7.17, L171/74-L171-99.

360 Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on*  
361 *Information Theory*, 13(1), 21–27. <http://doi.org/10.1109/TIT.1967.1053964>

362 Daumas, G. (2011). Non-electronic techniques to classify pig carcasses in small  
363 slaughterhouses. In *Second International Virtual Conference on pork Quality. Novembre,*  
364 *05to Decembre* (p. p 06-2001). Retrieved from  
365 [https://www.researchgate.net/profile/Gerard\\_Daumas/publication/285574394\\_NON-](https://www.researchgate.net/profile/Gerard_Daumas/publication/285574394_NON-ELECTRONIC_TECHNIQUES_TO_CLASSIFY_PIG_CARCASSES_IN_SMALL_SLAUGHTERHOUSES/links/565efc6308ae1ef92984212f.pdf)  
366 [ELECTRONIC\\_TECHNIQUES\\_TO\\_CLASSIFY\\_PIG\\_CARCASSES\\_IN\\_SMALL\\_SLA](https://www.researchgate.net/profile/Gerard_Daumas/publication/285574394_NON-ELECTRONIC_TECHNIQUES_TO_CLASSIFY_PIG_CARCASSES_IN_SMALL_SLAUGHTERHOUSES/links/565efc6308ae1ef92984212f.pdf)  
367 [UGHTERHOUSES/links/565efc6308ae1ef92984212f.pdf](https://www.researchgate.net/profile/Gerard_Daumas/publication/285574394_NON-ELECTRONIC_TECHNIQUES_TO_CLASSIFY_PIG_CARCASSES_IN_SMALL_SLAUGHTERHOUSES/links/565efc6308ae1ef92984212f.pdf)

368 Dirinck, P., Van Opstaele, F., & Vandendriessche, F. (1997). Flavour differences between  
369 northern and southern European cured hams. *Food Chemistry*, 59(4), 511–521.  
370 [http://doi.org/10.1016/S0308-8146\(97\)00012-5](http://doi.org/10.1016/S0308-8146(97)00012-5)

371 Edwards, D. B., Bates, R. O., & Osburn, W. N. (2003). Evaluation of Duroc- vs. Pietrain-sired  
372 pigs for carcass and meat quality measures. *Journal of Animal Science*, 81(8), 1895.  
373 <http://doi.org/10.2527/2003.8181895x>

374 Fàbrega, E., Gispert, M., Tibau, J., Hortós, M., Oliver, M. A., & Font i Furnols, M. (2011).  
375 Effect of housing system, slaughter weight and slaughter strategy on carcass and meat  
376 quality, sex organ development and androstenone and skatole levels in Duroc finished  
377 entire male pigs. *Meat Science*, 89(4), 434–439.

378 <http://doi.org/10.1016/j.meatsci.2011.05.009>

379 Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of*  
380 *Eugenics*, 7(2), 179–188. <http://doi.org/10.1111/j.1469-1809.1936.tb02137.x>

381 Font-i-Furnols, M., Čandek-Potokar, M., Daumas, G., Gispert, M., Judas, M., & Seynaeve, M.  
382 (2016). Comparison of national ZP equations for lean meat percentage assessment in  
383 SEUROP pig classification. *Meat Science*, 113, 1–8.  
384 <http://doi.org/10.1016/j.meatsci.2015.11.004>

385 Font-i-Furnols, M., Gispert, M., Soler, J., Diaz, M., Garcia-Regueiro, J. A., Diaz, I., & Pearce,  
386 M. C. (2012). Effect of vaccination against gonadotrophin-releasing factor on growth  
387 performance, carcass, meat and fat quality of male Duroc pigs for dry-cured ham  
388 production. *Meat Science*, 91(2), 148–154. <http://doi.org/10.1016/j.meatsci.2012.01.008>

389 Font i Furnols, M., & Gispert, M. (2009). Comparison of different devices for predicting the  
390 lean meat percentage of pig carcasses. *Meat Science*, 83(3), 443–446.  
391 <http://doi.org/10.1016/j.meatsci.2009.06.018>

392 Gispert, M., Àngels Oliver, M., Velarde, A., Suarez, P., Pérez, J., & Font i Furnols, M. (2010).  
393 Carcass and meat quality characteristics of immunocastrated male, surgically castrated  
394 male, entire male and female pigs. *Meat Science*, 85(4), 664–670.  
395 <http://doi.org/10.1016/j.meatsci.2010.03.021>

396 Gispert, M., Font i Furnols, M., Gil, M., Velarde, A., Diestre, A., Carrión, D., ... Plastow, G. S.  
397 (2007). Relationships between carcass quality parameters and genetic types. *Meat Science*,  
398 77(3), 397–404. <http://doi.org/10.1016/j.meatsci.2007.04.006>

399 Kempster, A. J., Chadwick, J. P., & Jones, D. W. (1985). An evaluation of the Hennessy  
400 grading probe and the SFK Fat-O-Meater for use in pig carcass classification and grading.  
401 *Animal Production*, 40(2), 323–329. <http://doi.org/10.1017/S0003356100025447>

402 Latorre, M. A., García-Belenguer, E., & Ariño, L. (2008). The effects of sex and slaughter  
403 weight on growth performance and carcass traits of pigs intended for dry-cured ham from  
404 Teruel (Spain). *Journal of Animal Science*, 86(8), 1933–1942.  
405 <http://doi.org/10.2527/jas.2007-0764>

406 Lucas, D., Brun, A., Gispert, M., Carabús, A., Soler, J., Tibau, J., & Font-i-Furnols, M. (2017).  
407 Relationship between pig carcass characteristics measured in live pigs or carcasses with  
408 Piglog, Fat-o-Meat'er and computed tomography. *Livestock Science*, 197, 88–95.  
409 <http://doi.org/10.1016/j.livsci.2017.01.010>

410 Nanni Costa, L., Stelletta, C., Cannizzo, C., Gianesella, M., Lo Fiego, D. P., & Morgante, M.



411 (2010). The use of thermography on the slaughter-line for the assessment of pork and raw  
412 ham quality. *Italian Journal of Animal Science*, 6(1s).  
413 <http://doi.org/10.4081/ijas.2007.1s.704>

414 Olsen, E. V., Candek-Potokar, M., Oksama, M., Kien, S., Lisiak, D., & Busk, H. (2007). On-  
415 line measurements in pig carcass classification: Repeatability and variation caused by the  
416 operator and the copy of instrument. *Meat Science*, 75(1), 29–38.  
417 <http://doi.org/10.1016/j.meatsci.2006.06.011>

418 Realini, C. E., Duran-Montgé, P., Lizardo, R., Gispert, M., Oliver, M. A., & Esteve-Garcia, E.  
419 (2010). Effect of source of dietary fat on pig performance, carcass characteristics and  
420 carcass fat content, distribution and fatty acid composition. *Meat Science*, 85(4), 606–612.  
421 <http://doi.org/10.1016/j.meatsci.2010.03.011>

422 Tous, N., Lizardo, R., Vilà, B., Gispert, M., Font-I-Furnols, M., & Esteve-Garcia, E. (2014).  
423 Effect of reducing dietary protein and lysine on growth performance, carcass  
424 characteristics, intramuscular fat, and fatty acid profile of finishing barrows 1. *J. Anim.*  
425 *Sci*, 92, 129–140. <http://doi.org/10.2527/jas2012-6222>

426 Vapnik, V. N., & Chervonenkis, A. (1964). A note on one class of perceptrons. *Automation and*  
427 *Remote Control*, 25(6), 937–945. Retrieved from  
428 [https://scholar.google.es/scholar?q=Vladimir+Vapnik+Alexey+Ya.+Chervonenkis&hl=ca](https://scholar.google.es/scholar?q=Vladimir+Vapnik+Alexey+Ya.+Chervonenkis&hl=ca&as_sdt=0%2C5&as_ylo=1960&as_yhi=1965)  
429 [&as\\_sdt=0%2C5&as\\_ylo=1960&as\\_yhi=1965](https://scholar.google.es/scholar?q=Vladimir+Vapnik+Alexey+Ya.+Chervonenkis&hl=ca&as_sdt=0%2C5&as_ylo=1960&as_yhi=1965)

430 Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., ...  
431 Whittington, F. M. (2008). Fat deposition, fatty acid composition and meat quality: A  
432 review. *Meat Science*. <http://doi.org/10.1016/j.meatsci.2007.07.019>

433 Wood, J. D., Enser, M., Whittington, F. M., Moncrieff, C. B., & Kempster, A. J. (1989). Wood  
434 et al., 1989 Backfat composition in pigs Differences between fat thickness groups and  
435 sexes..pdf. *Livestock Production Science Elsevier Science Publishers B.V.* Retrieved from  
436 [http://ac.els-cdn.com/0301622689900663/1-s2.0-0301622689900663-](http://ac.els-cdn.com/0301622689900663/1-s2.0-0301622689900663-main.pdf?_tid=02bf6b6e-300f-11e7-a6c9-00000aacb35d&acdnat=1493822848_b5a764336449f25b0d2b04164c239dad)  
437 [main.pdf?\\_tid=02bf6b6e-300f-11e7-a6c9-](http://ac.els-cdn.com/0301622689900663/1-s2.0-0301622689900663-main.pdf?_tid=02bf6b6e-300f-11e7-a6c9-00000aacb35d&acdnat=1493822848_b5a764336449f25b0d2b04164c239dad)  
438 [00000aacb35d&acdnat=1493822848\\_b5a764336449f25b0d2b04164c239dad](http://ac.els-cdn.com/0301622689900663/1-s2.0-0301622689900663-main.pdf?_tid=02bf6b6e-300f-11e7-a6c9-00000aacb35d&acdnat=1493822848_b5a764336449f25b0d2b04164c239dad)

439 Wood, J. D., Nute, G. R., Richardson, R. I., Whittington, F. M., Southwood, O., Plastow, G., ...  
440 Chang, K. C. (2004). Effects of breed, diet and muscle on fat deposition and eating quality  
441 in pigs. *Meat Science*, 67(4), 651–667. <http://doi.org/10.1016/j.meatsci.2004.01.007>

442