

Magnetic Resonance Imaging, texture analysis and regression techniques to non-destructively predict the quality characteristics of meat pieces

M.M. Avila, M.L. Durán, D. Caballero, T. Antequera, T. Palacios-Pérez, E. Cernadas and M. Fernández-Delgado

Version: accepted article

How to cite:

M.M. Avila, M.L. Durán, D. Caballero, T. Antequera, T. Palacios-Pérez, E. Cernadas and M. Fernández-Delgado (2019) Magnetic Resonance Imaging, texture analysis and regression techniques to non-destructively predict the quality characteristics of meat pieces. *Engineering Applications of Artificial Intelligence*, 82, 110 - 125.

Doi: <https://doi.org/10.1016/j.engappai.2019.03.026>

Copyright information:

©2019 Elsevier Ltd. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

Magnetic Resonance Imaging, texture analysis and regression
techniques to non-destructively predict the quality
characteristics of meat pieces

M.M. Ávila and M.L. Durán

School of Technology Universidad de Extremadura, Spain

D. Caballero and T. Antequera and T. Palacios-Pérez

Meat and Meat Products Research Institute, Universidad de Extremadura, Spain

Avda. de las Ciencias s/n, 10003, Cáceres Spain

E. Cernadas and M. Fernández-Delgado

Centro Singular de Investigación en Tecnoloxías da Información da USC (CiTIUS)

Universidade de Santiago de Compostela

Rúa Xenaro de la Fuente Domínguez, 15782, Santiago de Compostela, Spain.

E-mail: eva.cernadas@usc.es

February 25, 2019

Abstract

The quality of meat products is traditionally assessed by chemical or sensorial analysis, which are time consuming, need specialized technicians and destroy the products. The development of new technologies to monitor meat pieces using non-destructive methods in order to establish their quality is earning importance in the last years. An increasing number of studies have been carried out on meat pieces combining Magnetic Resonance Imaging (MRI), texture descriptors and regression techniques to predict several physico-chemical or sensorial attributes of the meat, mainly different types of pig ham and loins. In spite of the importance of the problem, the conclusions of these works are still preliminary because they only use the most classical texture descriptors and regressors instead of stronger methods, and because the methodology used to measure the performance is optimistic. In this work, we test a wide range of texture analysis techniques and regression methods using a realistic methodology to predict several physico-chemical and sensorial attributes of different meat pieces of Iberian

pigs. The texture descriptors include statistical techniques, like Haralick descriptors, local binary patterns, fractal features and frequencial descriptors, like Gabor or wavelet features. The regression techniques include linear regressors, neural networks, deep learning, support vector machines, regression trees, ensembles, boosting machines and random forests, among others. We developed experiments using 15 texture feature vectors, 28 regressors over 4 datasets of Iberian pig meat pieces to predict 39 physico-chemical and sensorial attributes, summarising 16,380 experiments. There is not any combination of texture vector and regressor which provides the best result for all attributes tested. Nevertheless, all these experiments provided the following conclusions: 1) the regressor performance, measured using the squared correlation (R^2), is from good to excellent (above 0.5625) for 29 out of 39 attributes tested; 2) the WAPE (Weighted Absolute Percent Error) is lower than 2% for 32 out of 37 attributes; 3) the dispersion in computer predictions around the true attributes is lower or similar than the dispersion in the labelling expert's for the majority of attributes (85%); and 4) differences between predicted and true values are not statistically significant for 29 out of 37 attributes using the Wilcoxon ranksum statistical test. We can conclude that these results provide a high reliability for an automatic system to predict the quality of meat pieces, which may operate on-line in the meat industries in the future.

1 Introduction

Hams and loins from Iberian pig, which is an autochthonous porcine breed developed traditionally in the SouthWest of Spain, are one of the most valuable meat products in this country. This is mainly ascribed to their exceptional sensorial attributes that depend on characteristics of raw material and processing conditions. Thus, not only characteristics of fresh pieces but also their modifications during the processing are important parameters to control the technological process of dry-cured hams and loins [1]. Temperature and relative humidity conditions during the processing lead to dehydration and, hence, to weight loss and a water activity decrease. Meat from Iberian pigs should contain plenty of intramuscular fat, which is an important characteristic, due to its positive influence on quality parameters on the final product, such as marbling, juiciness, odor, and aroma [2]. The determination of salt content is important from a microbiological point of view, but it also influences on the texture and flavor of the final product [3]. Color is also one of the most interesting characteristics of meat products [4], and for dry cured meat products it is the most relevant appearance property [5]. It is also important to study the final sensory quality of Iberian meat products, considering their most distinguished sensorial attributes, such as appearance, odor, taste and flavor [6]. Scientific studies on these meat products have carried out the sensory analysis objectively, with trained panellists and following standardised conditions.

Techniques usually carried out for determining these parameters related to the quality of meat

products are laborious and time consuming, they require the destruction of the pieces and a trained staff panel in the case of the sensory analysis. In this sense, MRI and computer vision techniques have been proposed as an alternative, since they are non-destructive, non-invasive, non-intrusive, non-ionizing and innocuous. The majority of works, that used MRI to determine quality characteristics of dry-cured products, are focused on hams, allowing to monitor the ripening process of Iberian [7], Parma [8], and S. Daniele hams [9] and to study Iberian hams as a function of pig feeding background [10, 11]. To our knowledge, only Cernadas et al. [12] and Caballero [13] analyzed loins by MRI, allowing: 1) an adequate product classification as a function of pig breeding; and 2) the prediction of sensory traits.

The general procedure in most of these studies takes three main steps: image acquisition, image analysis and data analysis. For image acquisition, high field MRI scanners have been used, which provide high quality images but they are expensive. A cheaper alternative are low field scanners, whose capability to acquire images of loin with predictive aims in food analysis, despite its low signal to noise ratio, has been demonstrated in recent works [13, 14].

Regarding the image analysis, most works [15, 13, 16, 17] have only used classical second order statistics like grey-level cooccurrence, grey-level run-length and neighboring gray-level dependence descriptors to extract texture information from the MRI slices. In relation to data analysis usual statistical tools, such as Pearson correlation coefficients or principal components analysis, have been used to process data from MRI and to determine quality parameters of products [8, 7, 9]. Perez-Palacios et al. [18, 15, 13] tested common regression techniques such as linear and isotonic regression to predict quality characteristics. However, these studies evaluated the regression methods following the well-known cross-validation methodology [19], including slices belonging to the same meat piece both in the training and testing sets. This methodology may provide optimistic results because the training set includes data from the same meat piece whose quality characteristics must be predicted.

Taking in mind these considerations, the current work is aimed to test: i) a wide variety of texture extraction techniques to analyse MRI; ii) a large amount of regression methods using a realistic evaluation methodology, with the final purpose of predicting different physico-chemical characteristics and sensorial attributes of meat products with high accuracy in a non-destructive way; and iii) a study focused on minimizing the needed images for a good prediction in order to improve the acquisition time.

The paper is organised as follows: Section 2 describes the materials used to obtain the MRI sequences and to develop the physico-chemical and sensory analysis on the meat pieces; Section 3 briefly explains the texture features extraction techniques and regression methods used; Section 4 shows the experimental setup (datasets and statistical validation) and Section 5 presents and discusses the results; and finally Section 6 summarises the main conclusions and proposes the

future work.

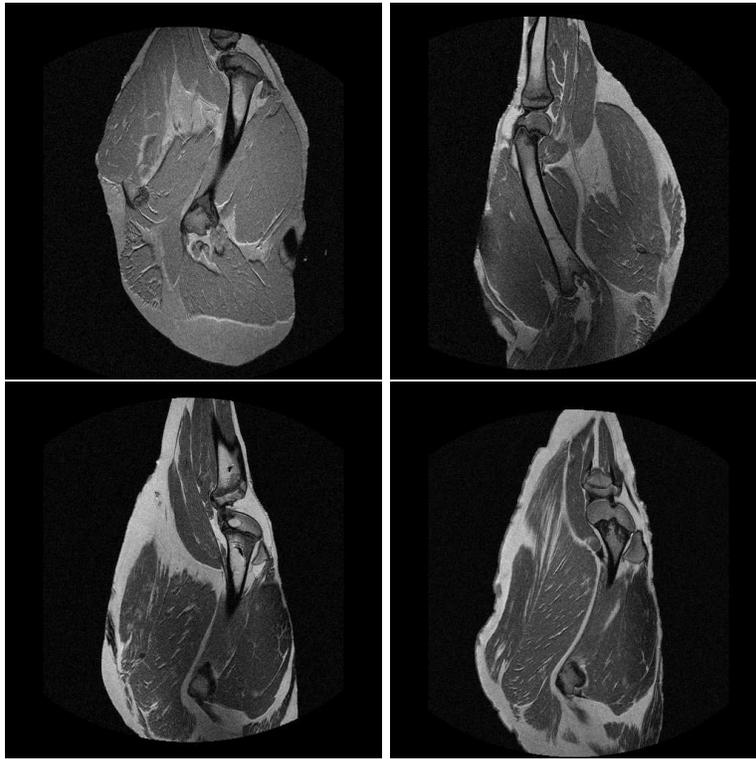


Figure 1: Examples of MRI images of hams at different stages of the processing: raw hams (upper left), end of post-salting (upper right), end of drying (lower left) and dry-cured hams (lower right).

2 Materials

Before describing the image acquisition technology and available data, it is important to emphasize that the physico-chemical and sensory analyses provide only one attribute value per loin/ham piece, although the MRI scanners give several images of each loin/ham piece.

2.1 MRI acquisition

Images of hams were firstly obtained using a medical purpose MRI scanner (Philips Gyroscan NT Intera 1.5 T) in the Infanta Cristina Hospital in Badajoz (Spain). This scanner was used with the quadrature whole-body coil, and sequences of T1 were applied with the following parameters: 120×85 cm 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), and 0.23×0.20 mm per pixel resolution. Sixty slices per ham piece were obtained. The MRI acquisition was done at a temperature of 20° C and it took approximately 28 minutes for each ham. All the images have

512 × 512 pixels and 256 grey levels. Figure 1 shows examples of Iberian ham at different stages of processing.

In relation to loins, they were scanned by using a low-field MRI scanner (ESAOTE VET-MR E-SCAN XQ 0.18 T) with a hand/wrist coil at the Animal Source Foodstuffs Innovation Services (SiPA) in the Faculty of Veterinary Science at University of Extremadura (Cáceres, Spain). The Iberian loins images were acquired with the spin echo (SE) T1-weighted sequence using the following default values: FOV, 150 × 150 mm; TE, 26 ms; slice thickness, 4 mm; TR, 630 ms; three acquisitions per sample. Twenty-nine slices per loin with 631 × 631 pixels and 256 grey levels were obtained. The MRI acquisition approximately took 50 minutes for each loin. Figure 2 shows examples of MRI images of loin.

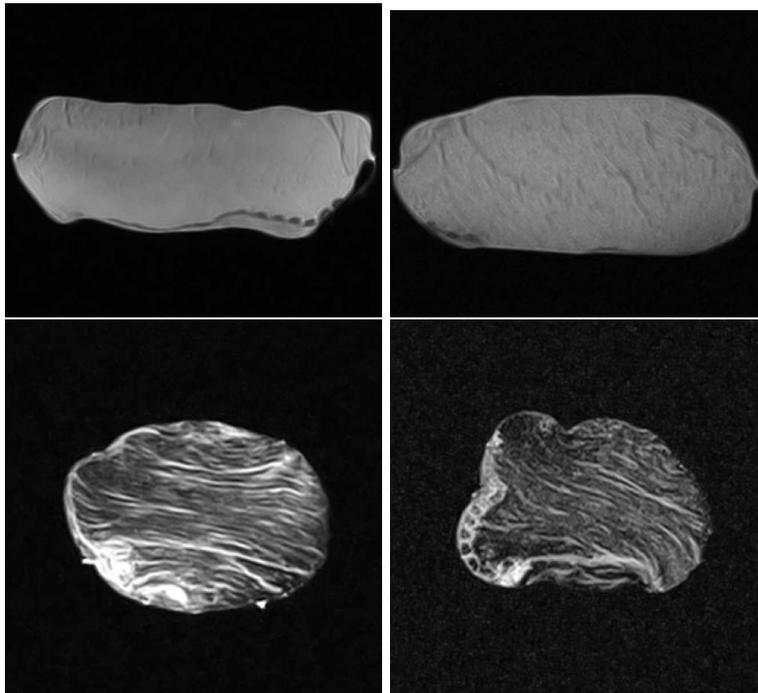


Figure 2: Examples of MRI images of Iberian fresh loin (upper images) and dry-cured loin (lower images).

This study has been carried out with four batches of different Iberian meat products: i) five fresh loins; ii) five dry-cured loins; iii) three hams at each stage of the processing: raw hams (0 days), end of post-salting (90 days), end of drying (270 days) and dry-cured hams (660 days); and iv) ten dry-cured hams.

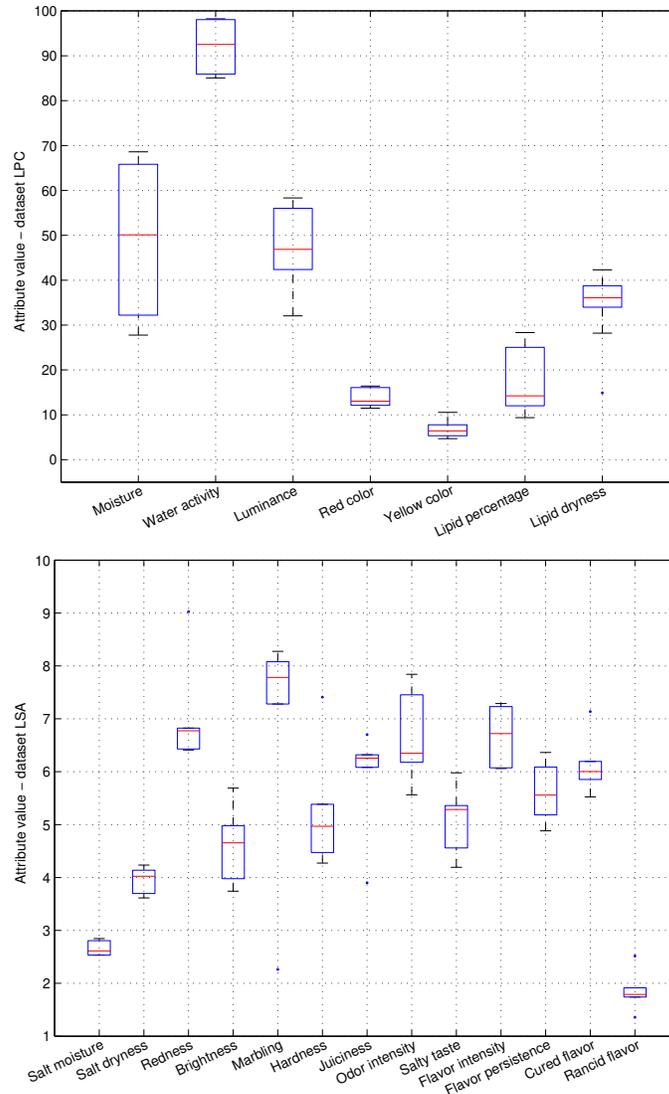


Figure 3: The upper and lower panels show, respectively, the box plots of the physico-chemical attributes of fresh and dry-cured loins (dataset LPC) and the sensorial attributes of dry-cured loins (dataset LSA).

2.2 Physico-chemical and sensory data

After the MR scanning, all the meat pieces have been destroyed, in order to calculate the true physico-chemical and sensorial attributes (which represent the ground truth) for each meat piece, which are used to validate the prediction methods. Fresh and dry cured loins were physico-chemically analysed by means of 7 attributes whose box plots, labeled as dataset LPC (Loin Physico-Chemical attributes), can be observed in the upper panel of Figure 3. The lower and upper ends of the box define the 25% and 75% quantiles, respectively, the red line is the median of

the values and the whiskers are the 0% and 100% quantiles. These 7 attributes are: moisture, water activity, instrumental color (boxes “luminance”, “red color” and “yellow color”, which correspond to L, a* and b* coordinates in the Lab color space) and lipid content (boxes “lipid percentage” and “lipid dryness”). The moisture was determined at 102 ± 2 °C by the official method (AOAC, 2000; reference 118 935.29). Water activity was measured in the system Lab Master-aw (NOVASINA AG, Switzerland) after calibration at 20-22 °C. Instrumental colors were measured using a Minolta CR-300 colorimeter (Minolta Camera 125 Corp., Meter Division. Ramsey, NJ). The lipid content of loins was determined gravimetrically with chloroform:methanol (2 : 1, *v/v*), according to the method described in [20]. Lipid content was calculated in wet and dry basis (lipid percentage and lipid dryness in Figure 3).

In addition, determination of salt content and sensory analysis were also carried out in dry-cured loins. Salt content was volumetrically analysed by the official method (AOAC, 2000; reference 971.19). A quantitative-descriptive analysis was applied for the sensory analysis, using a trained panel of thirteen members. The box plots with the attribute values, labeled as LSA (Loin Sensorial Attributes), can be observed in the lower panel of Figure 3, including two chemical attributes (salt moisture and dryness) and 11 sensorial attributes (traits) of Iberian dry-cured loins were assessed in a non-structured scale 0 – 10: redness, brightness, marbling, hardness and juiciness of lean; odor intensity; salty taste; flavor intensity and persistence; cured and rancid flavor. Analyses were developed as specified in [13].

The batches of hams at the different stages of processing and the dry-cured hams were analysed by means of moisture and lipid content, following the methodology previously described for loins (box plots of the values for the hams, labeled as H2, can be observed in the upper panel of Figure 4). Besides, a sensory analysis was developed, as described in [15], in the batch of dry-cured hams. In this case, 17 sensorial attributes were analysed: moisture (divided by 10 in the lower panel of Figure 4, in order to use the same scale as the other attributes); lipid content; lean redness, luminance (brightness), marbling, hardness, dryness, juiciness and pastiness; odor intensity; salty, sweet and bitter taste; flavor intensity and persistence; cured and rancid flavor. The box plots of these sensorial attributes for dry-cured hams (labeled as dataset H3) can be observed in the lower panel of Figure 4.

It is important to emphasize that although the established methods to assess the quality of Iberian meat products are the physico-chemical and sensorial analysis, both methods have an inherent experimental error. In order to give a true value for a physico-chemical or sensorial attribute, the experts repeat the measurement 3 or 13 times, respectively, and then they average the acquired values. We quantify this experimental error defining the TSTD (true standard deviation) as the average over the meat pieces of the standard deviation for each physico-chemical and sensorial attribute k :

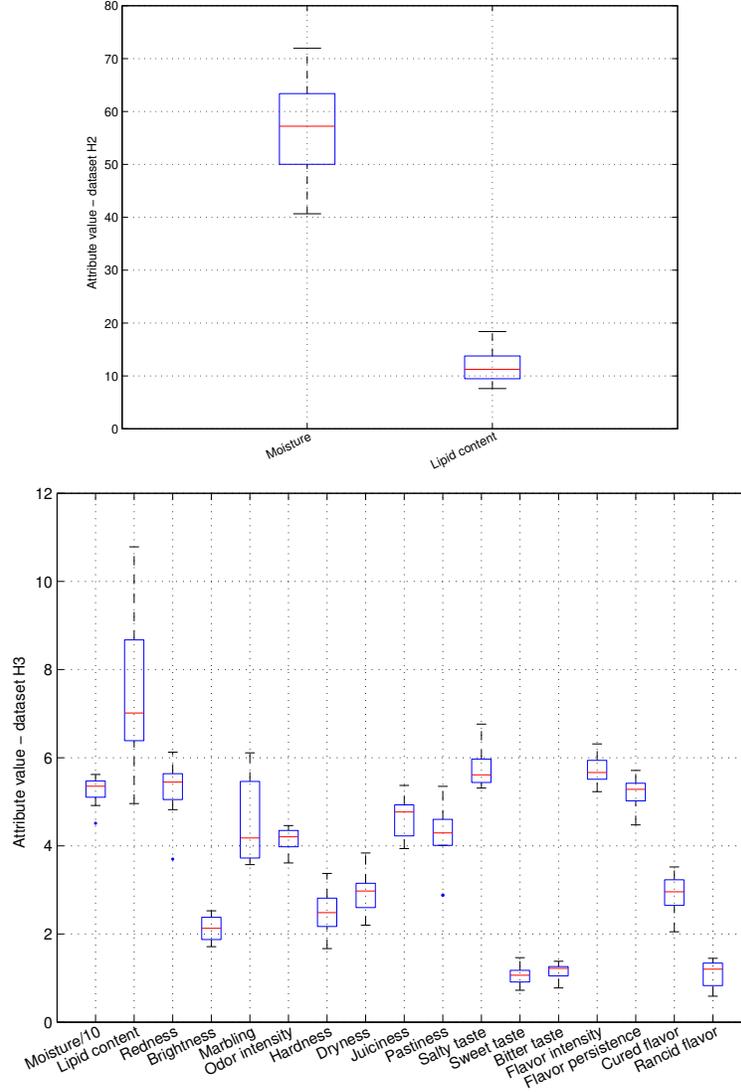


Figure 4: The upper and lower panels show, respectively, the box plots of the physico-chemical attributes of hams at different stages of the processing (dataset H2) and sensorial attributes of dry-cured hams (dataset H3).

$$TSTD_k = \frac{1}{N} \sum_{i=1}^N \sqrt{\frac{1}{M_k - 1} \sum_{j=1}^{M_k} (d_{ijk} - \bar{d}_{ik})^2} \quad (1)$$

where N is the number of meat pieces, M_k is the number of true measurements for attribute k , d_{ijk} is the j -th true measurement of meat piece i and attribute k , while \bar{d}_{ik} is the average over all the measurements for meat piece i and attribute k .

3 Methods

In this Section, we present a brief description of the texture features and regression techniques used in the experimental work. Many texture extraction methods must be applied to square regions of interest (ROIs), which must be selected from the original images. In the case of loin, a ROI of 128×128 was automatically extracted by thresholding the original image using the Otsu method [21]. Afterwards, the same combination of mathematical morphological filters are used in order to fill the holes in the segmented region and the ROI is extracted from the centroid of that region. For ham, the biceps muscle of each piece was extracted using the active contour algorithm proposed by Caro et al. [22]. Afterwards, for each slice of the biceps a squared 64×64 pixels ROI was extracted from the centroid. The ROIs were analysed by the different texture feature methods explained in the following subsection.

3.1 Texture features

We used a collection of 15 texture features included in the comparison developed by Cernadas et al. [23], which includes *statistical* techniques, like grey level cooccurrence matrix and sum and difference histograms (both second-order statistical methods), local binary patterns and fractal analysis; and *spectral* methods, like wavelets and Gabor filters. In the following, we describe briefly the texture feature vectors used:

- **Second-order statistical features.** The Grey Level Cooccurrence Matrix (GLCM) describes the probability of finding two pixels with the same value at different scales, or distances, and orientations, or angles [24, 25]. In many practical applications, the orientations used are 0° , 45° , 90° and 135° , and the scales range from one to eight pixels. The GLCM matrices are averaged for each scale and all orientations. Normally, some features are derived from GLCM matrices in order to reduce the high dimensionality and to be adequate as input patterns for regression methods. For each scale, we compute the contrast, homogeneity, correlation and energy of the GLCM matrix. We construct the vectors: **COMS**, which includes the four previous features for scale 1; and **MCOMS**, with the same four features for scales $\{1, 2, 3, 4, 5, 6, 7, 8\}$, summing 32 features. The Sum and Difference Histograms were introduced by Unser [26] as a faster and efficient alternative to GLCM for texture analysis. We use two vectors: **SDH**, which includes five features (energy, correlation, entropy, contrast and homogeneity) calculated for scale 1; and **MSDH** (multidimensional SDH), including 40 features, 5 for each scale in the above scale set.
- **Local binary patterns (LBP).** The LBP operator is a well-known texture analysis technique proposed by Ojala et. al. [27], which describes each pixel comparing its value with the

neighboring pixels. The main drawback of original method is the high dimensionality of the LBP descriptor, which has been partially addressed by some LBP versions [28, 27]. We use the following variants: 1) LBP uniform pattern [27], implemented in the Matlab language for scientific computing [29] by the LBP Matlab toolbox¹, which provides the following texture feature vectors: **LBP^{riu}** (LBP with uniform patterns), which uses $P = 8$ neighbors and radius $R = 1$, containing 10 features; and **MLBP^{riu}** (multi-resolution **LBP^{riu}**), which concatenates the features for $(P, R) \in \{(8, 1), (16, 2), (24, 3)\}$ summing 54 features; 2) the Local Binary Count (LBC), with the following feature vectors [30]: **CLBC** (Completed LBC) and **MCLBC** (multi-resolution Completed LBC), with 9 and 53 features respectively. The vector **CLBC** use $P = 8$ and radius $R = 1$, while **MCLBC** concatenates the features for $(P, R) \in \{(8, 1), (16, 2), (24, 3)\}$. We use the publicly available² LBC implementation.

- **Fractal texture features.** The fractal dimension is also used for texture analysis [31, 32]. Based on our previous experience [33], we use the probability method [34] computing the fractal dimension for the box sizes $L = \{3, 5, 7, 9, 11, 13, 15, 17, 19, 21\}$ pixels, yielding a vector of 10 features called **MFP** (MultiFractal Probability). Xu et al. [35] introduced a texture descriptor called **MFS** (MultiFractal Spectrum), based on fractal geometry theory and box-counting methods, with 26 features. We use the publicly available MFS³ code with its default parameter values.
- **Wavelet features.** Discrete wavelet transform (DWT) representation is a theory for multi-dimensional signal decomposition [36, 37] which recursively apply filters to decompose the image into low-pass and high-pass frequency bands. In order to obtain a compact representation useful for texture analysis, the mean and variance of the energy distribution for the transformed coefficients in each sub-band and decomposition level are used. We compute two feature vectors calculating the mean and variance of the energy over the two analysis filters used (Haar and Daubechies). In both cases we use 3 levels of decomposition and take the statistics only over the low-low decomposition sub-band. With Daubechies filters we use a filter of size 4. The feature vectors are: **Haar** (Haar filter); and **Daub4** (Daubechies filtering), both vectors with 20 features. For the calculation of wavelet texture features, we use the **Wavelab** toolbox⁴.

Nick G. Kingsbury [38] proposed the dual-tree complex wavelet transform (DT-CWT), which some authors consider as a special case of Gabor Filters with complex coefficients. Celik et al. [39] uses as texture features the variance and entropy of the magnitude of the coefficients.

¹<http://www.cse.oulu.fi/CMV/Downloads/LBP Matlab>

²<http://home.ustc.edu.cn/~zyknight/CLBC.rar>

³<http://www.cfar.umd.edu/~fer/website-texture/texture.htm>

⁴http://www-stat.stanford.edu/~wavelab/Wavelab_850/index_wavelab850.html

For comparative purposes with the DWT, we use 3 levels of decomposition and compute two feature vectors: **DTCWT** (mean and variance statistics for 3 scales and 6 directions, 36 features) and **DTCWT-VH** (variance and entropy statistics for 3 scales and 6 directions, 36 features). For the calculation of DT-CWT, we use the Matlab toolbox⁵ provided by the authors.

- **Gabor filters.** This is a very popular technique for texture classification [40], which consists of sinusoidal waves modulated by a Gaussian envelope. The Gabor filter parameters used are: central frequency of the filter at the highest frequency ($F_m = 0.327$), number of frequencies ($n_F = 4$), number of orientations ($n_O = 6$), the frequency ratio is half-octave frequency spacing ($F_r = \sqrt{2}$) and standard deviation of the Gaussian envelope, i.e. the *smoothing* parameters ($\gamma = \eta = 0.5$), as it is recommended by Bianconi and Fernández [41], assuming uniform angular spacing. We use the Simplegabor toolbox⁶ for multiresolution Gabor filtering of 2-D signals (images). Given a bank of digital Gabor filters $G_{ij}(x, y)$ with $i \in \{1, \dots, n_F\}$ and $j \in \{1, \dots, n_O\}$, the Gabor transform of the input image is computed for each filter of the bank. We use the Gabor with Normalisation for Illumination invariant, **GNI** texture feature vector with 48 features, which also includes the mean μ_{ij} and the standard deviation σ_{ij} of the magnitude of each transformed image.

3.2 Regression methods

A collection composed by the best 28 regressors achieved in the comparative study of Sirsat [42, 43, 44] was applied in the present work (see the list below). The majority of them (24 regressors) are implemented in the R language for statistical computing [45], using packages detailed in the list below. Additionally, we also execute other 4 popular regression methods which are implemented in other platforms: 1) deep learning neural network (which we name dlkeras in the regressor list below), using the library *Keras* with the Theano interface in the Python programming language; 2) support vector regression, using the LibSVM library accessed via its C++ interface (named *svr*); 3) generalised regression neural network (named *grnn*) included in the Matlab neural network toolbox; and 4) extreme learning machine with Gaussian kernels (named *kelm*), also programmed in Matlab.

Most regressors in our collection have tunable hyper-parameters, i.e., parameters which must be specified previously to training, whose values often have a strong influence on the regressor performance. In these cases, it is a good practice to try several values for each hyper-parameter in a trial-and-error procedure, and to select the value which provides the best performance on

⁵<http://eeweb.poly.edu/iselesni/WaveletSoftware/index.html>

⁶<http://www2.it.lut.fi/project/simplegabor/downloads/src/simplegabortb>

the available data. In order to optimise this performance for each regressor, its tunable hyper-parameters and the list of values for each hyper-parameter should be known. For the regressors which are not implemented in R, we directly specify the list of tunable hyper-parameters and the values used for tuning, which are specified by the regressor documentation. On the other hand, for the regressors implemented in R we used the list of hyper-parameter values provided by the R package “Classification and Regression Training” (**caret**). Specifically, this package [46] provides the `getModelInfo` function, which returns for each regressor and dataset a list of recommended values which should be used for tuning each hyper-parameter (a list of the available models and their hyper-parameters can be found in this link⁷). This utility avoids the need to analyze the documentation of every regressor in order to know proper values to be used for the hyper-parameter tuning. The values used for each regressor are specified in the following list. The notation **a:b:c** means a list of values from **a** to **c** with step **b** (where the step is missing, its value is assumed to be 1).

1. **lm** is the linear regression provided by the **stats** R package [47], which performs multivariate linear regression.
2. **penalized** is the penalized linear regression (**penalized** package), which is regularized by weighting two penalties: L1, also called least absolute shrinkage and selection operator (LASSO), is the sum of the absolute values of the linear regression coefficients; and L2, also called ridge penalty, is the sum of squared coefficients. The weights of both penalties are tunable hyper-parameters (λ_1 and λ_2 arguments in the **penalized** R function) with values $\{2^i\}_0^4$ and $\{2^i\}_0^3$, respectively [48].
3. **krlsRadial** is the regularized least squares regression (**KRLS** package) with Gaussian radial basis function kernel [49]. The regularization parameter ($\lambda=0.1$) specifies the trade-off between model fit and complexity, and the only tunable hyper-parameter is the kernel spread (σ), with 10 values in the set $\{10^i\}_{-7}^2$.
4. **foba** [50] develops ridge regression with forward, backward and sparse input selection (**foba** package). The hyper-parameters are the regularization (λ) for ridge regression, with 10 values in the range 10^{-5} -0.1, and the number of selected inputs (**k**) for prediction, with values 2 and the number of inputs.
5. **avNNet** is the model averaged neural network (**caret** package), a committee of 5 neural networks of the same size trained using different random seeds, whose attributes are averaged. The hyper-parameters are the network size, with 7 values, which depends on the specific data, and the weight decay, with three values 0, $\{10^{-i}\}_2^4$.

⁷<https://topepo.github.io/caret/train-models-by-tag.html>

6. **grnn** is the generalized regression neural network [51] implemented by the Matlab neural network toolbox. The Gaussian spread is a hyper-parameter, tuned with 14 values in the range 0.001–2, which governs the smoothness of the approximation (higher for small spread values).
7. **kelm** is the extreme learning machine (ELM) neural network with Gaussian kernel [52] using the publicly available Matlab code⁸. The hyper-parameters are the regularization parameter C and the kernel spread tuned with values $\{2^i\}_{-5}^{14}$ and $\{2^i\}_{-16}^8$, respectively.
8. **dlkeras** is the deep learning neural network [53, 54] implemented by the `Keras` module [55], of the Python programming language. This network has three hidden layers tuned with 25, 50 and 75 neurons for each layer (27 combinations).
9. **svr** is the epsilon-support vector regression with Gaussian kernel, using the C++ interface to the LibSVM library [56]. The regularization hyper-parameter C and the kernel spread γ are tuned with the same values as `kelm` (see above).
10. **M5** is a regression tree [57] implemented by the Weka Data Mining Software⁹ and accessed from a R program through the `RWeka` package, tuning three flags: `pruned` and `smoothed`, with values `yes` and `no` each one; and `rules/trees`, a flag to select between a tree of a rule set.
11. **cubist** [58] is a M5 rule-based regressor with corrections based on nearest neighbors in the training set (`Cubist` package). Its hyper-parameters are the number of training committees (5 odd values between 1 and 10) and the number of neighbors for prediction (5 values from 2 to 20).
12. **earth** [59] is the multivariate adaptive regression spline (MARS), implemented by the `earth` package. The maximum number of terms in the model (`nprune`) is tuned with up to 15 data-dependent values between 2 and 17.
13. **bagEarth** is a bagging ensemble of multivariate adaptive regression splines (MARS) base regressors implemented by the `caret` and `earth` packages. The only hyper-parameter is the maximum number of terms (`nprune`) in the pruned regression model, with 10 data-dependent values.
14. **gbm** is the generalized boosting regression model (`gbm` package), called stochastic gradient boosting machine in the `caret` model list (footnote 7 above). The hyper-parameters are the

⁸<http://www.extreme-learning-machines.org>

⁹<https://www.cs.waikato.ac.nz/ml/weka>

maximum depth of input interactions (`interaction.depth`), with values 1:5, and the number of trees for prediction (`n.trees`), with values 50:50:250. We use a Gaussian distribution and `shrinkage=0.1`.

15. **gamboost** (`mboost` package) is a boosting ensemble of generalized additive models [60]. The number of initial boosting iterations (`mstop`) is tuned with 10 linearly spaced values between 50 and 500.
16. **rf** is the random forest [61] ensemble of averaged random regression trees (`randomForest` package). The number of inputs selected at each tree (`mtry`) is tuned with 10 values between 2 and the number of inputs.
17. **Boruta** [62] is the random forest ensemble with additional feature selection (`Boruta` package). The only hyper-parameter is `mtry`, tuned as `rf`.
18. **RRF** is the regularized random forest (`RRF` package), which uses regularization to select inputs in random forest. The hyper-parameters are `mtry`, with 3 values between 2 and the number of inputs, and the regularization and importance coefficients (`coefReg` and `CoefImp`), with 2 and 3 values between 0 and 1, respectively.
19. **cforest** (`party` package) is a random forest ensemble of conditional inference trees [61], each one fitting one bootstrap sample. The only hyper-parameter is `mtry`, tuned as `rf`.
20. **extraTrees** [63] is the ensemble of extremely randomized regression trees (`extraTrees` package). Its tunable hyper-parameters are `mtry` (tuned as `rf`) and the minimum sample size to split a node (`numRandomCuts`), with values 1:10.
21. **qrf** is the quantile regression forest (`quantregForest` package), a tree-based ensemble which generalizes random forest in order to estimate conditional quantile functions. The `mtry` parameter is tuned as `rf`. The quantile prediction threshold (`what` argument in the `predict.quantregForest` function) is set to 0.5.
22. **rqlasso** develops quantile regression with least absolute shrinkage and selector operator (LASSO) penalty, using the `rq.lasso.fit` function in the `rqPen` package. This method fits a quantile regression model with the LASSO penalty [64], tuning the regularization hyper-parameter (λ), with 10 values from 0.1 to 10^{-4} .
23. **brnn** [65] is the Bayesian regularized neural network (`brnn` package), which uses inference to determine the weights of the squared error and the squared sum of unnormalized network weights [66]. The number of hidden neurons is tuned with 15 values from 1 to 15.

24. **bartMachine** [67] is the Bayesian additive regression tree (**bartMachine** package). The number of trees is set to 50 and the tunable hyper-parameters are the prior boundary (**k**), with 3 data-dependent values between 2 and 5, and the base value in tree prior to decide if a node is terminal or not (**α**), with values 0.9, 0.945 and 0.99.
25. **gaussprPoly** develops Gaussian process regression with polynomial kernel (**polydot**), tuning the kernel hyper-parameters **degree** and **scale**, both with three values in the ranges 1:3 and $\{10^{-i}\}_1^3$, respectively.
26. **lars** [68] is the least angle regression (**Lars** package). The **lasso** type and **fraction** mode are specified for training and prediction respectively, and the fraction hyper-parameter (called **s**) is tuned with values from 0.05 to 1.
27. **ppr** [69] performs the projection pursuit regression (**stats** package), which iteratively calculates the coefficients that minimize the fraction of unexplained variance which is explained by each function. The only hyper-parameter is the number of terms to be included in the final model (**nterms**), with values 1:10.
28. **enet** is the elasticnet regression model (**elasticnet** package), computed using the least angle regression - elasticnet (LARS-EN) algorithm [70]. There are two hyper-parameters (5 values each one): the quadratic penalty or regularization hyper-parameter (λ , with values 0, $\{10^{-i}\}_1^4$) and the fraction (**s**) of the L1 norm of the coefficient vector relative to the norm at the full least squares solution (the **fraction** mode is used in the **predict.enet** function, with values 0.05, 0.28, 0.52, 0.76, 1).

4 Experimental setup

This Section describes the datasets and the evaluation methodology used in the experimental work.

4.1 Dataset description

Four datasets were formed depending on the physico-chemical and sensorial attributes available for each meat piece (our ground truth): LPC, LSA, H2 and H3. Each dataset was constructed with computational (from MRI), physico-chemical and/or sensory data. The **LPC** (Loin Physico-Chemical) dataset contains 5 fresh loins and 5 dry-cured loins of Iberian pigs, specifically values of features from 290 images (29 images per loin) and from the following physico-chemical attributes: moisture; water activity; L, a^* , b^* ; lipid content in wet and dry basis. There are 30 values for each

physico-chemical characteristic (3 values per loin) and 60 values for color components (6 values per loin).

The **LSA** (Loin Sensorial Attributes) dataset contains the five dry-cured loins included in LPC dataset, including features from 145 images and salt in wet and dry basis (3 values per loin); redness, brightness, marbling, hardness and juiciness of lean; odor intensity; salty taste; flavor intensity and persistence; cured and rancid flavor (12 values per loin).

The **H2** (ham during maturation process) dataset contains data from 12 tights of Iberian pigs at four stages of maturation process (three for each stage). The number of slices for each tight varies with the piece and maturation stage. In order to form a dataset with the same number of patterns per piece, 16 slices of each piece were considered. So, this dataset contains data of features from 192 images (12 hams \times 16 slices) and from the physico-chemical attributes moisture and lipid content in wet basis (2 values per ham).

The **H3** (cured ham) dataset contains data from 10 dry-cured Iberian ham, specifically data from 150 images (15 slices \times 10 pieces) and from the following physico-chemical characteristics and sensorial attributes: moisture and lipid content in wet basis (1 value per ham); redness, luminance, marbling, hardness, dryness, juiciness and pastiness of lean; odor intensity; salty, sweet and bitter taste; flavor intensity and persistence, cured and rancid flavor (13 values per ham).

Combinations of the 15 texture feature vectors and the 28 regressors were tested in each dataset in order to predict each quality attribute. Thus, the number of experiments developed for dataset LPC is 2940, which corresponds to 15 texture vectors times 28 regressors times 7 attributes to be predicted. For the rest of datasets, we have $15 \times 28 \times 13$ attributes = 5460 experiments for dataset LSA, 840 (= $15 \times 28 \times 2$ attributes) for H2 and 7140 (= $15 \times 28 \times 17$ attributes) for H3.

4.2 Evaluation methodology

In order to evaluate the regressor performance for the prediction of physico-chemical and sensorial attributes, a variant of the common cross-validation methodology has been used to develop the tuning of regressor hyper-parameters and to guarantee a realistic evaluation. The main changes introduced by this experimental methodology are: 1) three data partitions (or sets), for training, validation and test, are used instead of classical cross-validation, which only uses training and test sets; and 2) the dataset is created leaving one meat piece out, instead of the usual random partitioning of the image collection, i.e., the images of one meat piece are used as test set, and the images of the remaining meat pieces are randomly divided into training and validation sets. The texture feature vector calculated for each image is used as input pattern for the regressors, so there are so many patterns as images.

The validation set is used to fix the tunable hyper-parameters. If two datasets (training and test) were only used, the test set would be the only available to evaluate: 1) the model performance

with each hyper-parameter value in order to select its best value and 2) the performance of the final model with the selected hyper-parameter value. However, this methodology would optimistically bias the performance evaluation, because it would measure the final performance on the same test set where the model, trained with the selected value of the hyper-parameter, achieved its best performance. Thus, the performance of the final model for other datasets, different to the test set, would be expected to be lower. The need to select an “optimal value” for the tunable hyper-parameters requires to evaluate the model, trained for each hyper-parameter value, in a dataset different to the test set, where the performance of the final model, trained with the best hyper-parameter values, will be tested. Hence the need of validation sets.

Dataset	N_i	n_i	#images	#Patterns			#attributes
				#training	#validation	#testing	
LPC	10	29	290	188	73	29	7
LSA	5	29	145	94	22	29	13
H2	12	16	192	124	52	16	2
H3	10	15	150	97	38	15	17

Table 1: Second, third and fourth columns are respectively the number of meat pieces (N_i) and images per piece (n_i) and the total number of images of dataset $i = \{LPC, LSA, H2, H3\}$. The following columns are the number of patterns used in the training, validation and test sets, and the last column is the number of attributes to be predicted for each dataset.

To develop experiments, N_i partitions were created, being N_i the number of meat pieces in the dataset i , with $i = \{LPC, LSA, H2, H3\}$. Therefore, there are N_i training, validation and test partitions. For each partition j , with $j = 1, \dots, N_i$, the patterns P_{ij} belonging to the i -th meat piece are used as test set, and the rest of patterns (which belongs to images of meat pieces $l \neq i$) are randomly distributed into the training set (65% of the patterns) and validation set (35%). All the inputs and attributes are pre-processed (standardized) in order to have zero mean and standard deviation one. Table 1 summarizes the number of training, validation and test patterns for each dataset. Each regressor is trained on the N_i training partitions for each combination of its hyper-parameter values (regressors may have zero, one, two or three tunable hyper-parameters), and tested on its corresponding validation partitions. For each combination of hyper-parameter values, we select for testing the combination with the lowest RMSE (Root Mean Squared Error) defined by:

$$RMSE = \sqrt{\frac{1}{N_V} \sum_{k=1}^{N_V} (y_k - t_k)^2} \quad (2)$$

where N_V is the number of validation patterns, y_k is the attribute value predicted by the regressor and t_k is the true value (ground truth) of this attribute for the k -th validation pattern. For each combination of hyper-parameter values, the average RMSE over the N_i validation sets is calculated, and the combination with the lowest RMSE is selected for testing. Finally, the regressor is trained, using this selected combination of its hyper-parameter values, on the N_i training partitions, and the final performance is averaged over the N_i test partitions. Instead of using the RMSE as the final performance measure, we use the squared correlation coefficient R^2 because it provides an absolute measurement of the regressor quality, since its values are bounded between 0 and 1. The value of R^2 is given by:

$$R^2 = \frac{\left(\sum_{k=1}^{N_T} (y_k - \bar{y})(t_k - \bar{t}) \right)^2}{\sum_{k=1}^{N_T} (y_k - \bar{y})^2 \sum_{k=1}^{N_T} (t_k - \bar{t})^2} \quad (3)$$

where N_T is the number of test patterns, while \bar{y} and \bar{t} are the average values of predicted and true attributes, respectively, over the N_T test patterns. However, since all the patterns of the same test partition share the same output value (because they belong to the same meat piece, which obviously has only one value for each physico-chemical or sensorial attribute), R^2 can not be computed for each test partition separately, because in eq. 3 we would have $t_k = \bar{t}$, for $k = 1, \dots, N_T$, so that $\sum_{k=1}^{N_T} (t_k - \bar{t}) = 0$ in the denominator. In order to avoid this drawback, we calculate R^2 considering jointly all the test partitions, so that in eq. 3 we set $N_T = N_i n_i$, i. e. the number N_i (with $i = \{LPC, LSA, H2, H3\}$) of test partitions multiplied by the number n_i of patterns in each test partition, while y_k and t_k are the predicted and true attribute values, respectively, for all the test patterns in all the test partitions; and \bar{y} and \bar{t} are the averages of y_k and t_k over all the test partitions. This is the R^2 value calculated considering each image as a pattern, which will be named in the following as “Image R^2 ”.

Additionally, for a practical application we are interested in the prediction of a unique value for each meat piece, and not for each image of the meat piece. The predicted value for a meat piece might be calculated from the predicted values for their images taking their average value or, even better, their median value, which is a more robust statistical descriptor. Therefore, we additionally calculate the “Piece R^2 ” as the squared correlation coefficient between the true value for each meat piece and the predicted value, also for each piece, calculated as the median of the predicted values over all the images of the same meat piece.

The MAE (Mean Absolute Error) and WAPE (Weighted Absolute Percentage Error) can also measure the accuracy of a prediction. The MAE is the average absolute difference between the predicted and true values of the attribute, while WAPE measures the same difference but expressed

as a percentage of the attribute mean. Both measures are defined for each physico-chemical or sensorial attribute k by:

$$MAE_k = \frac{1}{N} \sum_{i=1}^N |y_{ik} - t_{ik}| \quad WAPE_k(\%) = \frac{100 \sum_{i=1}^N |y_{ik} - t_{ik}|}{\sum_{i=1}^N t_{ik}} \quad (4)$$

where y_{ik} and t_{ik} are the predicted and true values for attribute k , while N is the number of samples. As we mentioned, both MAE_k and $WAPE_k$ can be calculated at the image level, in which case N represents the number n_i of RMI slices, with $i = \{LPC, LSA, H2, H3\}$, or at the meat piece level, in which case N represents the number N_i of meat pieces.

Dataset	Attribute	Feature	Regressor	Image R^2	Piece R^2
LPC	Moisture	mfs	RRF	0.97399	0.98321
	Water activity	lbp	brnn	0.98670	0.99319
	Luminance	sdhc	brnn	0.83754	0.89144
	Redness	mfs	Boruta	0.79281	0.82514
	Yellowness	gni	rqlasso	0.49286	0.55110
	Lipid content	mlbp	gamboost	0.63743	0.75074
	Lipid dryness	gni	svr	0.33121	0.39543
LSA	Salt moisture	daub4	svr	0.77203	0.96250
	Salt dryness	msdhc	krIsRadial	0.92589	0.97021
	Redness	mfs	M5	0.54318	0.48924
	Luminance	dtcwtVH	RRF	0.70834	0.92933
	Marbling	dtcwt	extraTrees	0.44261	0.61367
	Hardness	haar	earth	0.78153	0.88443
	Juiciness	dtcwt	extraTrees	0.55494	0.66212
	Odor intensity	mfp	bartMachine	0.80395	0.88277
	Salty taste	dtcwt	cubist	0.50672	0.64080
	Flavor intensity	dtcwt	rf	0.87716	0.97233
	Flavor persistence	mfp	bartMachine	0.91344	0.93700
	Cured flavor	msdhc	ppr	0.55838	0.74970
	Rancid flavor	mfs	earth	0.62910	0.68779

Table 2: Best squared correlations (R^2) achieved by some combination of texture feature vector and regressor for the different attributes of datasets LPC and LSA.

The Wilcoxon signed rank-sum test [71] will be used to check whether predicted values and true measurements of each attribute are different. This test compares two samples and tests the null hypothesis that both belong to distributions with the same mean. When the p -value provided by this test is lower than 0.05 the null hypothesis should be rejected, which means that

both samples do not belong to distributions with the same mean, i.e., difference between them is statistically significant. When $p \geq 0.05$, the null hypothesis can not be rejected, so the difference is not statistically significant.

5 Results and discussion

Dataset	Attribute	Feature	Regressor	Image R^2	Piece R^2
H2	Moisture	msdhc	bagEarth	0.87763	0.89956
	Lipid content	sdhc	bagEarth	0.80222	0.82825
H3	Moisture	gni	gaussprPoly	0.66551	0.82557
	Lipid content	dtcwtVH	bagEarth	0.57891	0.71907
	Redness	clbc	gaussprPoly	0.90337	0.91918
	Luminance	lbp	rqlasso	0.83185	0.84785
	Marbling	gni	brnn	0.54744	0.65563
	Odor intensity	dtcwt	foba	0.46149	0.52423
	Hardness	mfp	dlkeras	0.53792	0.77298
	Dryness	gni	krlsRadial	0.39372	0.54720
	Juiciness	dtcwtVH	krlsRadial	0.12706	0.21411
	Pastiness	msdhc	foba	0.31010	0.38175
	Salty taste	lbp	penalized	0.91023	0.91562
	Sweet taste	dtcwtVH	bartMachine	0.57272	0.67130
	Bitter taste	mclbc	krlsRadial	0.88146	0.88157
	Flavor intensity	dtcwtVH	earth	0.28571	0.22666
	Flavor persistence	mclbc	penalized	0.24586	0.34631
	Cured flavor	mclbc	krlsRadial	0.40708	0.50060
Rancid flavor	coms	svr	0.78085	0.92066	

Table 3: Best squared image and piece correlations (R^2) achieved by some feature vector and regressor for the different attributes of datasets H2 and H3.

Tables 2 and 3 show the combination of texture feature vectors and regressors which achieve the highest squared correlation (R^2) to predict each quality attribute of loin (LPC and LSA datasets) and ham (H2 y H3 datasets) meat pieces, respectively. Classical definition [72] for the correlation intervals and their significance considers that a R^2 value between 0 and 0.0225 means that the two vectors under comparison (in our case, these vectors are the true and predicted values, t_k and y_k , for each MRI slice) are not correlated at all; R^2 between 0.0225 and 0.25 means bad to moderate correlation between them; R^2 between 0.25 and 0.5625 means moderate to good correlation; and $R^2 > 0.5625$ means very good to excellent correlation. Taking this criterion in mind and analysing Tables 2 and 3 (column “Image R^2 ”), the values of R^2 are very good to excellent for 22 out to 39

attributes, moderate to good for 15 attributes and bad to moderate for only 2 attributes (juiciness and flavor persistence of dry-cured ham, H3). Considering meat piece (column “Piece R^2 ”), the R^2 values increase in relation to the image R^2 being good to excellent for 29 attributes, moderate to good for 8 attributes and only bad to moderate for 2 attributes (juiciness and flavor intensity of H3).

Pos.	Regr.-Feat.	Rank	Pos.	Regr.-Feat.	Rank	Pos.	Regr.-Feat.	Rank
1	bartMach-mfp	107.55	11	svr-mfs	131.15	21	RRF-mfp	139.55
2	svr-daub4	114.75	12	krlsRad-mfs	131.25	22	Boruta-mfp	139.60
3	extraTrees-dtcwt	121.95	13	svr-mcoms	131.30	23	extraTrees-daub4	140.10
4	krlsRad-daub4	123.10	14	dlkeras-mfp	132.35	24	extraTrees-mfp	140.40
5	rf-dtcwt	123.70	15	extraTrees-clbc	135.15	25	cforest-mfp	140.75
6	Boruta-dtcwt	123.90	16	RRF-dtcwt	135.80	26	RRF-dtcwtVH	141.15
7	extraTrees-dtcwtVH	124.90	17	extraTrees-mlbp	136.35	27	Boruta-dtcwtVH	141.50
8	svr-haar	126.20	18	bartMach-dtcwt	137.10	28	extraTrees-lbp	142.20
9	extraTrees-mclbc	126.80	19	Boruta-mclbc	137.75	29	rf-dtcwtVH	142.50
10	krlsRad-haar	130.80	20	rf-mfp	139.40	30	bartMach-dtcwtVH	144.30

Table 4: Table with the Friedman ranks of the 30 best combinations of texture feature and regressor for predicting the physico-chemical and sensorial attributes of loin (LPC and LSA).

Since the best combination of texture feature and regressor is not the same for all the attributes, we use the Friedman ranking [73] of the squared correlation (R^2) values over the physico-chemical and sensorial attributes of loins and hams in a separated way, in order to find the combination of texture feature and regressor which achieves the best R^2 over all the loin or ham attributes. Table 4 shows, in decreasing order, the 30 best combinations of texture features and regressor for all the physico-chemical and sensorial attributes of loin.

The regressor **bartMachine** combined with the multifractal probability texture feature (vector MFP) is in position 1 with a Friedman rank of 107.55, which means that this regressor is approximately in the position 107 in average over all the combinations of feature vectors and regressors (420 combinations). The high value (107.55) of this rank means that this combination regressor + feature, which is globally the best, is still far from being the most appropriate combination for most attributes (which would require a much lower rank, e.g. between 1 and 10). The next seven positions are for wavelet features (vectors Daub4, dtcwt, dtcwtVH and haar) with different regressors: svr (ranks 114.75 and 126.20 with vectors Daub4 and dtcwt, respectively), extraTrees (121.95 with dtcwt), krlsRad (123.10 with dtcwt), rf (123.60 with dtcwt) and Boruta (123.90 with dtcwt). This conclusion can also be extracted from the upper panel of Figure 5, which compares the R^2 achieved by the best combination of texture feature + regressor for each attribute (in

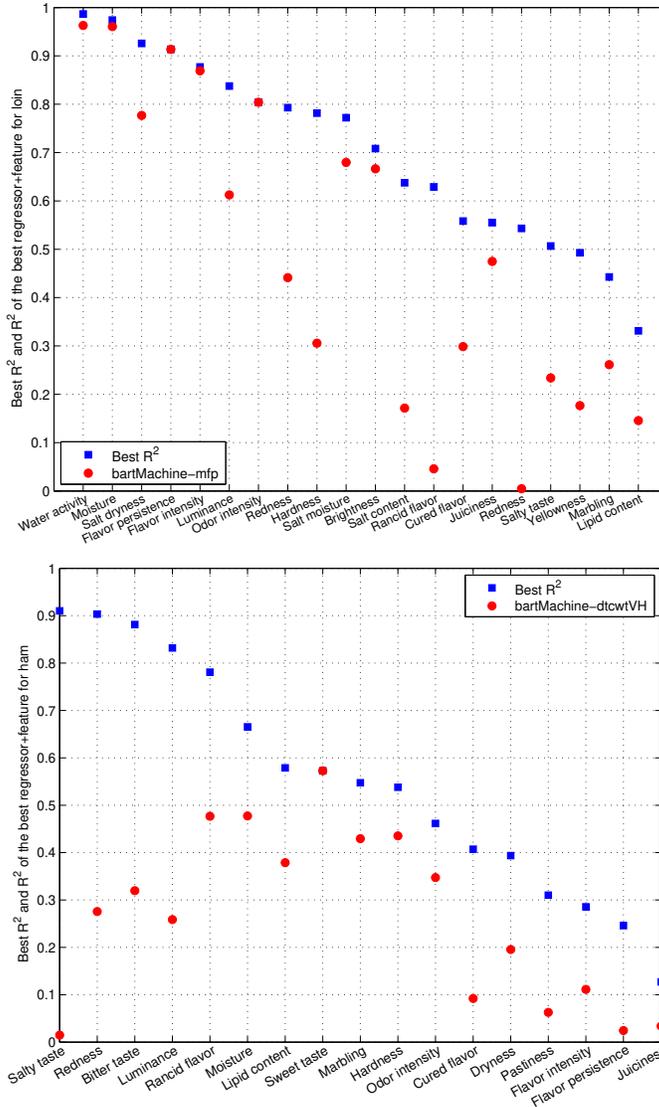


Figure 5: Comparison of R^2 achieved by the best combination of texture feature + regressor for each attribute (in blue) and by the globally best combination over all the attributes (in red) for loin (LPC and LSA datasets, upper panel) and ham (H3 dataset, lower panel).

blue), and the R^2 achieved by the combination with the lowest Friedman rank (in red). The best combination for loin (bartMachine + MFP) is near the best R^2 (difference below 0.1) for 8 out of 20 attributes, and difference is middle (0.1-0.3) for other 6 attributes, overcoming 0.3 for the remaining 6 attributes. The deep learning network (dlkeras) achieves the 14th position with MFS feature in this ranking, achieving the best result with MFP feature in the Hardness attribute (see Table 3).

The datasets of ham meat pieces (H2 and H3) are very different: H2 dataset encloses meat

Pos.	Regr.+Feat.	Rank	Pos.	Regr.+Feat.	Rank	Pos.	Regr.+Feat.	Rank
1	bartMach+dtcwtVH	89.41	11	krlsRad+dtcwtVH	128.35	21	Boruta+dtcwt	137.94
2	bartMach+dtcwt	92.94	12	rf+dtcwtVH	130.82	22	grnn+dtcwt	138.94
3	bartMach+gni	106.71	13	cforest+dtcwt	131.12	23	kelm+dtcwt	141.88
4	rf+gni	118.35	14	RRF+dtcwt	131.18	24	krlsRad+daub4	142.12
5	cforest+gni	121.47	15	RRF+dtcwtVH	131.82	25	pnlz+dtcwtVH	143.18
6	RRF+gni	122.76	16	Boruta+gni	132.82	26	gambst+dtcwtVH	143.88
7	extraTrees+gni	123.71	17	svr+dtcwtVH	133.76	27	brnn+dtcwtVH	144.00
8	krlsRad+gni	124.53	18	bartMach+mfp	134.82	28	kelm+gni	144.65
9	extraTrees+dtcwt	128.00	19	cforest+dtcwtVH	134.88	29	gbm+dtcwtVH	145.00
10	rf+dtcwt	128.29	20	kelm+dtcwtVH	136.18	30	grnn+dtcwtVH	146.65

Table 5: Table with the 30 best combinations of textures features and regressor for predicting the physico-chemical and sensorial attributes of dataset H3 (ham) using the Friedman rank.

pieces during the ripening process, while H3 dataset only includes dry-cured hams. So, these datasets are not studied together. For dataset H2, bagEarth (bagging ensemble of MARS base regressors) combined with the textures features Sum and Difference Histograms (vectors SDHC and MSDHC) achieves the best R^2 for both physico-chemical attributes (see column “Image R^2 ” in Table 3): 0.88 with MSDHC for moisture (0.87 with SDHC, not included in the table) and 0.80 with SDHC for lipid content (0.77 with MSDHC, not included in the table). In consequence, bagEarth combined with SDHC is useful to predict both attributes.

Table 5 shows the best combinations of texture features and regressor for all the physico-chemical and sensorial attributes of dry-cured ham (dataset H3) according to the Friedman rank. The regressor **bartMachine** combined with wavelet texture features (vectors dtcwtVH and dtcwt) are the best with ranks of 89.4 and 92.9, respectively. The bartMachine also achieves the third position combined with another frequencial texture descriptor, the Gabor texture feature (vector GNI). The following regressors are different variants of random forest (regressors rf, cforest, RRF and extraTrees) combined with GNI, whose Friedman ranks are 118, 121, 122 and 123, respectively. In every case, the first positions are for frequency texture descriptors (vectors DTCWT-VH, DTCWT and GNI). In the lower panel of Figure 5 the best combination (bartMachine + dtcwt-VH, in red) is below the best R^2 (in blue) by less than 0.1 in only 2 of 17 attributes, between 0.1-0.3 for 9 attributes and above 0.3 for the remaining 6 attributes.

Considering jointly the combination regressor-feature which achieves the best R^2 for each attribute (Tables 2 and 3) and the best position in the Friedman rankings (Tables 4 and 5) for loin and ham, some regressors outstand: bartMachine (which achieves the best R^2 in 3 attributes and positions 1 and 1-3 in loin and ham rankings), svr (3 attributes, positions 2 and 17), extraTrees

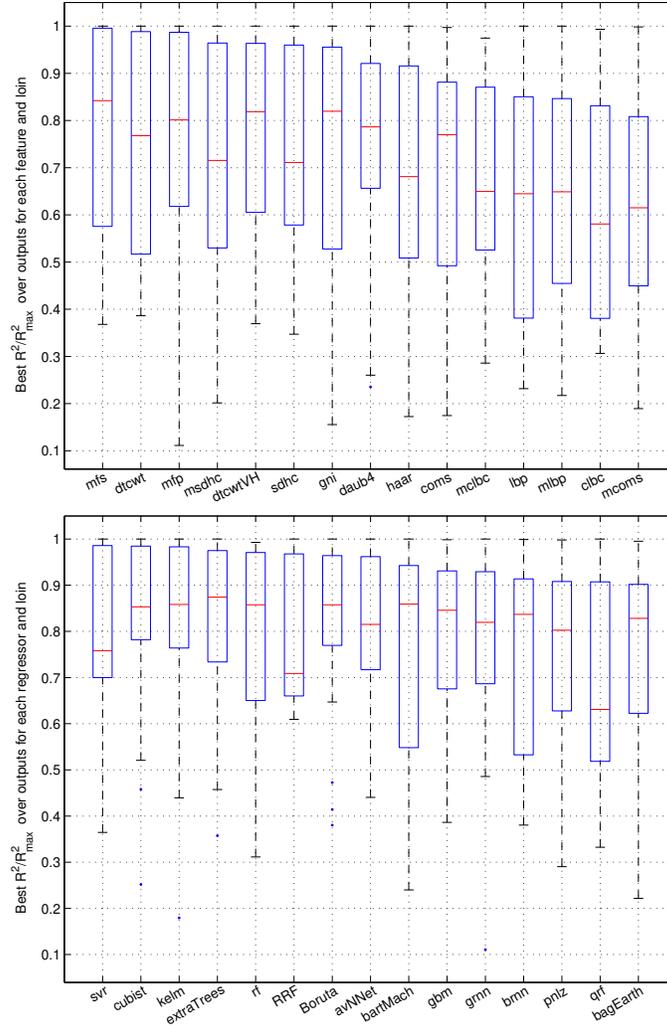


Figure 6: Upper panel: box plots of the best R^2 achieved by some regressor, divided by the globally best R^2 (named R^2_{\max}), over the attributes of loin (datasets LPC and LSA) for each feature. Lower panel: box plots of the best R^2 achieved by some feature, divided by R^2_{\max} , over the attributes of loin (same datasets) for each regressor.

(2 attributes, positions 3 and 7), krlsRadial (4 attributes, positions 4 and 8) and rf (1 attribute, positions 5 and 4). Other well-performing regressors are earth and bagEarth (3 attributes each one). These regressors also achieve very good results in a recent and very extensive comparison of regression methods [44], although the globally best regressors in this comparison (cubist and M5) only win in the current paper for 1 attribute each one. This is not surprising, because we can not expect that a given method will be the best in all the problems. Similarly, the deep learning network (dlkeras) only achieves the best R^2 in the current study for one attribute (hardness of cured ham, H3) and it is not ranked in the first positions. This also happens in [44], where dlkeras

achieves the 18th position in a ranking of 77 regressors, despite of the good performance of deep networks in many problems.

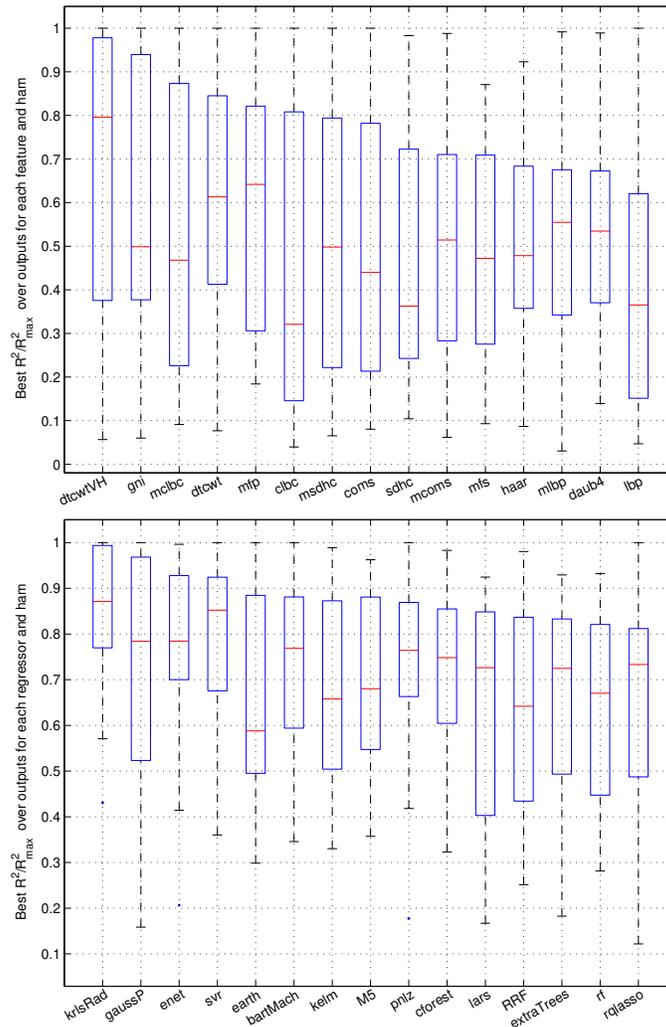


Figure 7: Upper panel: box plots of the best R^2 achieved by some regressor, divided by the globally best R^2 , over the attributes of ham (dataset H3), for each feature. Lower panel: box plots of the best R^2 achieved by some feature, divided by R_{\max}^2 , over the attributes of ham (dataset H3) for each regressor.

Figures 6 and 7 show the boxplots of R^2/R_{\max}^2 for texture features and regressors. The upper panels show one box for each texture feature, labelled in the horizontal axis. For example, the box of feature mfs shows the values of R^2/R_{\max}^2 achieved by this feature for all the attributes: for each attribute, R^2 is the best squared correlation achieved by this feature with any regressor, and R_{\max}^2 is the best R^2 achieved by any feature and regressor (also for that attribute). The lower panels show one box for each of the best 15 regressors, labelled by the horizontal axis. Each box (e.g.,

for regressor svr) plots the values of R^2/R_{\max}^2 achieved by this regressor over all the attributes. For each attribute, R^2 is the best value achieved by that regressor over all the features, and R_{\max}^2 is the best R^2 achieved by any regressor and feature for that attribute. In both panels, boxes are sorted by decreasing upper end of the box, which identifies the 75% quantile of the distribution. Considering features and loins (upper panel of Figure 6), multi-fractal (mfs and mfp) and wavelet (dtcwt) exhibit the highest upper ends, very near to one, while the following four features (msdhc, dtcwtVH, sdhc and gni) are about 0.95 and daub4 and haar are about 0.92. The remaining features are below 0.9. With respect to regressors (lower panel of Figure 6), the best 15 regressors exhibit 75% quantile above 0.9, being the best ones svr, cubist (with the narrowest box), kelm, extraTrees (which has the highest median), rf and RRF. Considering features and ham (upper panel of Figure 7), only wavelet (dtcwtVH, which exhibits clearly the highest 75% quantile and the highest median) and Gabor (gni), both frequencial features, have upper ends above 0.9, while the following four features (mclbc, dtcwt, mfp and clbc) lie between 0.8 and 0.9, and the remaining ones are below 0.8. Considering regressors (lower panel of Figure 7), krlsRadial exhibits the best median, 75% quantile and the narrowest box, followed by gaussprPoly, enet and svr (with upper end above 0.9), while the remaining regressors have upper ends below 0.9.

5.1 Results validation in meat industry

Table 6 shows the quality measures WAPE, MAE and TSTD for the combination of texture feature vector and regressor which best predicts each physico-chemical and sensorial attribute of all datasets. Globally, the WAPE is less than 1% for 20 out of 37 attributes, between 1% and 2% for 12 attributes and higher than 2% for only 5 attributes (2.58% for redness, 4.19% for marbling, 2.64% for hardness and 2.99% for rancid flavor in dataset LSA; and 2.12% for rancid flavor in dataset H3). The MAE evaluates the mean dispersion of the computer prediction values around the attribute, which is the average over the true measurements, and the TSTD evaluates the mean dispersion of the true measurements. We can see in Table 6 that this dispersion is lower for computer prediction than for true measurements in all sensorial attributes (datasets LSA and H3), except for marbling in dataset LSA. For most physico-chemical attributes (all the LPC attributes excepting luminance and redness, and the H2 attributes), the dispersions are higher for computer predictions than for true values. The dispersion in the true values might be influenced by the low number of measurements for physico-chemical and sensorial attributes (3 and 13 values, respectively). The values in column p -value are provided by a Wilcoxon test whose null hypothesis is that computer predictions and true measurements come from distributions with equal means. The column Sign shows the statistical significance of difference between predicted and true values: a value of YES means that this difference is significant, so the null hypothesis is rejected (i.e., both values do not come from distributions with equal means), which happens when $p < 0.05$; a

Set	output	features	regressor	WAPE (%)	MAE	TSTD	p -value	Sign.
LPC	Moisture	mfs	RRF	0.35	1.7007	0.9339	0.91171	NO
	Water activity	lbp	brnn	0.04	0.0038	0.0020	0.88826	NO
	Luminance	sdhc	brnn	0.49	2.3478	2.4713	0.99372	NO
	Redness	mfs	Boruta	0.53	0.7373	0.8514	0.31737	NO
	Yellowness	gni	rqlasso	1.34	0.9106	0.6196	0.50672	NO
	Lipid content	mlbp	gamboost	1.57	2.7024	1.1606	0.28378	NO
	Lipid dryness	gni	svr	1.58	5.4370	3.2442	0.00318	YES
LSA	Salt moisture	daub4	svr	0.39	0.0521	0.1048	0.90097	NO
	Salt dryness	sdhc	krlsRadial	0.26	0.0519	0.1318	0.74002	NO
	Redness	mfs	M5	2.58	0.9163	1.2729	0.00069	YES
	Luminance	dtcwtVH	RRF	1.09	0.2516	2.2234	0.83695	NO
	Marbling	dtcwt	extraTrees	4.19	1.4118	1.3683	0.01413	YES
	Hardness	haar	earth	2.64	0.6987	2.0244	0.08731	NO
	Juiciness	dtcwt	extraTrees	1.66	0.4867	1.8682	0.50611	NO
	Odor intensity	mfp	bartMachine	1.07	0.3579	1.6413	0.13950	NO
	Salty taste	dtcwt	cubist	1.45	0.3671	1.9371	0.03237	YES
	Flavor intensity	dtcwt	rf	0.52	0.1746	1.8331	0.12398	NO
	Flavor persistence	mfp	bartMachine	0.59	0.1648	2.4031	0.54743	NO
	Cured flavor	msdhc	ppr	1.44	0.4411	2.1485	0.00159	YES
	Rancid flavor	mfs	earth	2.99	0.2785	2.1733	0.00000	YES
H2	Moisture	msdhc	bagEarth	0.41	2.8218	0.6924	0.87710	NO
	Lipid content	sdhc	bagEarth	0.85	1.2208	0.6170	0.71819	NO
H3	Redness	clbc	gaussprPoly	0.76	0.4046	1.5916	0.25505	NO
	Luminance	lbp	rqlasso	0.75	0.1587	1.3805	0.55594	NO
	Marbling	gni	brnn	0.98	0.4412	1.8699	0.78878	NO
	Odor intensity	dtcwt	foba	0.43	0.1777	1.8850	0.66347	NO
	Hardness	mfp	dlkeras	0.88	0.2212	1.6833	0.16533	NO
	Dryness	gni	krlsRadial	0.98	0.2871	2.1488	0.01651	YES
	Juiciness	dtcwtVH	krlsRadial	0.71	0.3318	2.2363	0.93065	NO
	Pastiness	msdhc	foba	1.20	0.4985	2.3763	0.87929	NO
	Salty taste	lbp	penalized	0.47	0.2682	1.2453	0.74699	NO
	Sweet taste	dtcwtVH	bartMachine	1.06	0.1134	1.0339	0.04791	YES
	Bitter taste	mclbc	krlsRadial	1.09	0.1255	1.3411	0.00634	YES
	Flavor intensity	dtcwtVH	earth	0.40	0.2308	1.4931	0.93879	NO
	Flavor persistence	mclbc	penalized	0.43	0.2251	2.1561	0.09481	NO
	Cured flavor	mclbc	krlsRadial	1.09	0.3135	1.7110	0.33841	NO
	Rancid flavor	coms	svr	2.12	0.2328	0.8870	0.20846	NO

Table 6: The quality measures WAPE (%), MAE, TSTD, p -value and significance of the Wilcoxon test for all datasets and attributes (see details in the text).

value of NO means that the null hypothesis can not be rejected (i.e., difference is not statistically significant) because $p \geq 0.05$. In order to apply the test, both vectors must be of the same length, but the number of true measurements depends on the attribute and varies from 3 until 13, so we select the same number of predicted values from the slices located in the center of MRI sequences. The physico-chemical attributes (moisture and lipid content) of dataset H3 are not considered in this analysis because there is only one true measurement of each attribute, so the number of attributes is $39-2=37$. According to these p -values, difference between computer predictions and true values (column Sign. in Table) is not significant for 28 out of 37 attributes. For 6 of the remaining 9 attributes (salty taste, cured flavor and rancid flavor in dataset LSA; dryness, sweet taste and bitter taste in dataset H3), difference is statistically significant, but MAE is lower than TSTD, which means that the dispersion of predictions around the true value is lower than the dispersion of the expert’s scores.

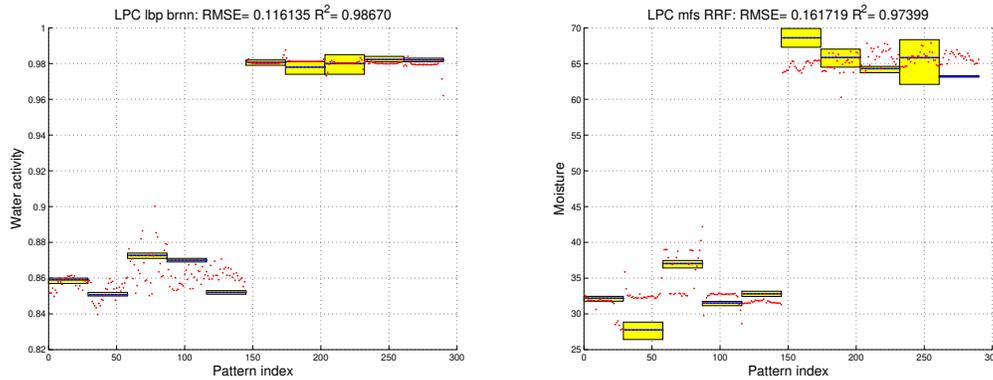


Figure 8: Predicted and true values (red and blue points, respectively) of all MRI slices for the attributes: water activity (left) and moisture (right) of dataset LPC. Yellow rectangles show the dispersion of the true measurements.

Figures 8 and 9 show the predicted and true values for the best combination of texture vector and regressor for the physico-chemical attributes water activity (for which the best texture feature vector is LBP and the best regressor is brnn) and moisture (mfs vector and RRF regressor) of dataset LPC (Figure 8) and the sensorial attribute marbling for datasets LSA (dtcwt and extraTrees) and H3 (gni and brnn) in Figure 9, to illustrate the meaning of the results showed in Table 6. The blue and red points are respectively the true and predicted values for each attribute. As mentioned, the blue points are constants within each interval, because they are the mean of values over the expert board for sensorial attributes, and over different trails for physico-chemical attributes. The vertical limits of each yellow rectangles show the interval of the true measurements for that meat piece. For the physico-chemical attributes of LPC (water activity and moisture, see

Figure 8), the R^2 is very high (0.986 for water activity and 0.973 for moisture, see the panel titles), the WAPE is very low (0.04% and 0.35%, see the first two lines of Table 6) and the predicted and true distributions are not statistically different (p -value > 0.05), so the prediction is very good. Nevertheless, the MAE is always higher than TSTD for both attributes (0.0038-0.0020 for water activity, and 1.70-0.93 for moisture), which means that the dispersion of predicted values is higher than for the true measurements, so many red points in Figure 8 are outside the yellow squares. This situation is repeated in other physico-chemical attributes.

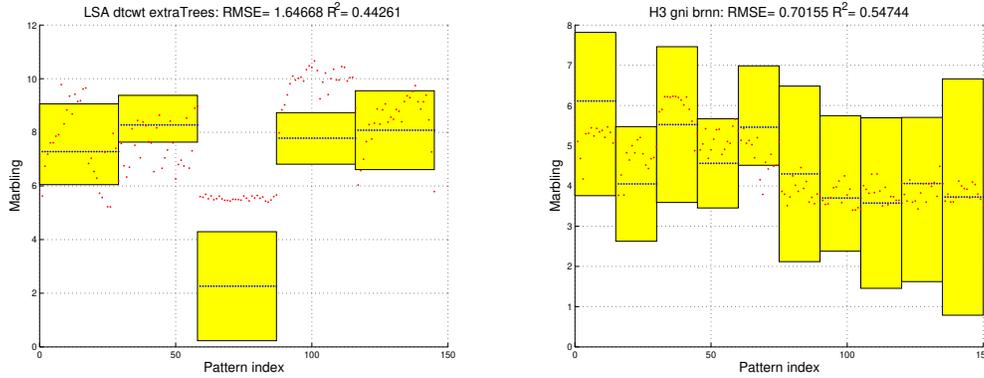


Figure 9: Predicted and true values (red and blue points, respectively) of all MRI slices for the attribute marbling of datasets: LSA (left) and H3 (right). Yellow rectangles show the dispersion of the true measurements.

As examples of the opposite case are the sensorial output marbling in datasets LSA and H3 (Figure 9), with lower image R^2 (0.44 and 0.55, respectively, see the panel titles), which means good to moderate correlation. Nevertheless, the WAPE of marbling is much higher for dataset LSA (4.19%) than for H3 (0.98%), see Table 6, and difference between predicted and true values is statistically significant in LSA, but not in H3. Consequently, the red points are always inside the yellow squares for dataset H3 (right panel in Figure 9), while many predictions (red points) fall outside the yellow squares for dataset LSA (left panel in Figure 9). The case of marbling in dataset H3 is very frequent for all sensorial attributes, in which there are high variability among experts and then, the predicted value is always within the limits of expert's scores (i.e., the red points are inside yellow squares for all meat pieces). So, in spite of the R^2 is not so high in some cases (e.g. cured flavor in dataset H3, with $R^2 = 0.40$ and juiciness, with $R^2 = 0.12$), sometimes with p -value < 0.05 , the computer prediction is reliable enough. For the majority of cases, the WAPE is lower than 2% (see Table 6), the MAE is lower than TSTD and difference between predicted and true values is not statistically significant. The situation of marbling for dataset LSA (low image and piece R^2 but MAE $<$ TSTD with $p < 0.05$) only happens for redness in dataset LSA.

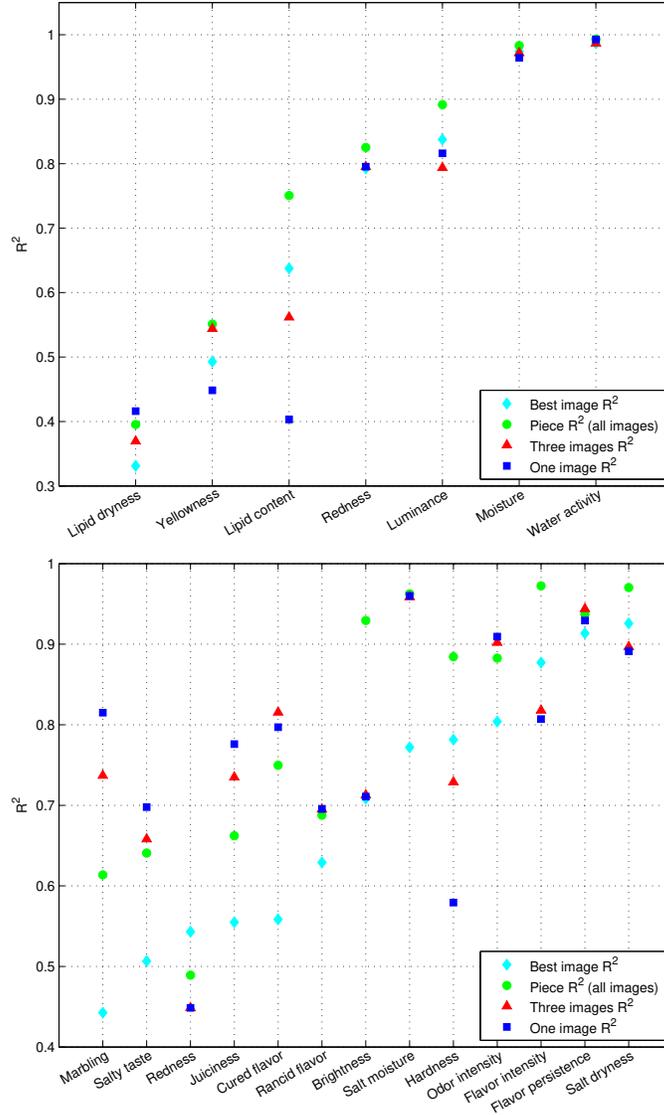


Figure 10: Image R^2 and piece R^2 using one, three and all images of a MRI sequence to predict all physico-chemical and sensorial attributes of the datasets LPC (upper panel) and LSA (lower panel).

In summary, the use of the best combination of texture feature and regression technique for each quality characteristic provides the following quality measures: i) the R^2 is from good to excellent for 29 out of 39 attributes tested, as reported by the last column of Tables 2 and 3; ii) the WAPE is below 2% for 32 out of 37 attributes, as reported by Table 6; iii) the dispersions in computer predictions around the true value and the dispersion in true measurements (columns MAE and TSTD in Table 6) are comparable, or even lower, for the majority of attributes; and iv) according to a Wilcoxon test, the dispersion (MAE) of the computer predictions is statistically

lower or equal than the variability of the measurement (TSTD) for 34 out of 37 attributes (see the last two columns in Table 6), being $MAE > TSTD$ for only 3 of 37 attributes.

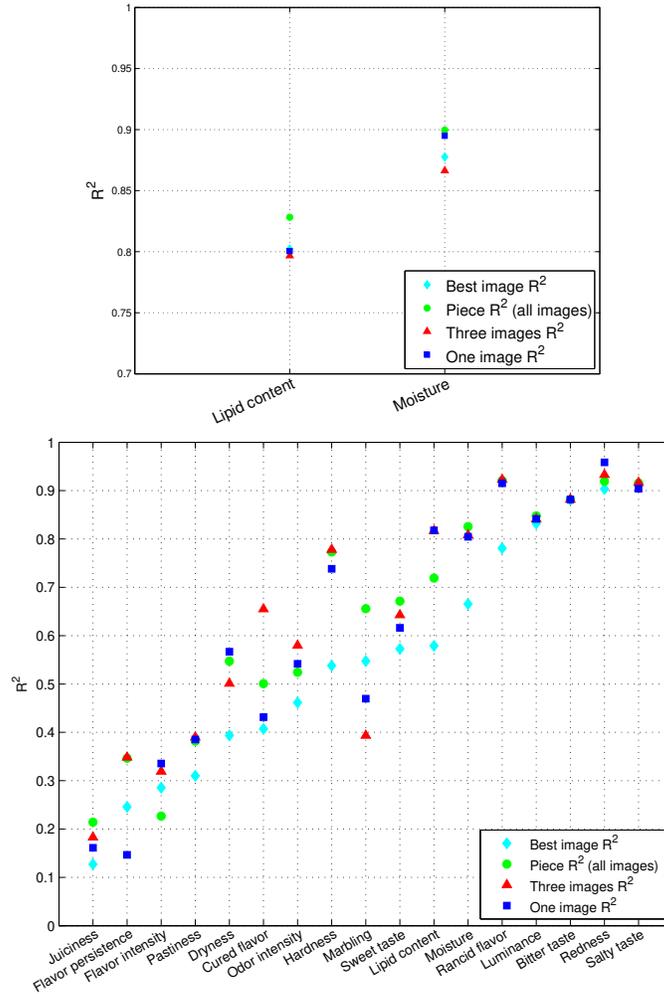


Figure 11: Image R^2 and piece R^2 using one, three and all images of a MRI sequence to predict all physico-chemical and sensorial attributes of the datasets H2 (upper panel) and H3 (lower panel).

These results confirm the high reliability of the use of MRI, texture analysis and regressors to predict physico-chemical and sensorial attributes of loins and ham pieces. Nevertheless, the main drawback in order to develop an automatic quality assessment system to operate in a meat industry is the high scanning time needed to obtain the MRI sequence (about 50 minutes to scan 30 slices). So, it would be desirable that the system develops the prediction based on fewer MR images, in order to decrease this prediction time. Figures 10 and 11 show: i) the best R^2 calculated at image level (legend “Best image R^2 ”, which corresponds to column Image R^2 in Tables 2 and 3); ii) the R^2 calculated at piece level using all slices of MRI sequence, with legend “Piece R^2 (all images)”; iii) the three central slices (legend “Three images R^2 ”); and iv) the central slice (legend

“One image R^2 ”). The piece R^2 (green circle), calculated using all images of its MR sequence, overcomes the best image R^2 (cyan diamond) for all datasets and attributes, except for redness in LPC dataset and flavor intensity in H3 dataset with differences less than 0.1. If we try to predict the attributes of the piece using three images R^2 (red triangle) or one image R^2 (blue square), they overcome the best image R^2 for 31 of 39 attributes, being below the piece R^2 (green circle) for half of attributes. The main conclusion is that the majority of physico-chemical and sensorial attributes of the meat pieces can be predicted with high fiability using only 3 MR slices (or even one for some attributes) of a meat piece, which implies a strong decreasing of scanning time. This fact makes this technology more feasible to implement in the meat industries.

6 Conclusions and future work

The adquisition of Magnetic Resonance Imaging (MRI) sequences of meat pieces allows to monitor them in a non-destructive and innocuous way, avoiding the use of chemical and sensorial analysis, which is very time consuming, expensive and requires specialized technicians. We develop an exhaustive comparison of 15 texture feature vectors selected from the statistical and frequency descriptors (including second order statistics, local binary patterns, Gabor and wavelet filters, among others) and 28 regression techniques (including most of the state of art methods as the Gaussian kernel support vector regression and extreme learning machine, deep learning, LASSO, random forest and elastic net, among others) in order to predict 7 physico-chemical and 17 sensorial attributes on four batches (LPC, LSA, H2 and H3) of different Iberian meat pieces (loins and hams). The physico-chemical attributes predicted are: moisture, water activity, lipid percentage and dryness, and instrumental color (luminance, red color and yellow color). The sensorial attributes are: lean redness, brightness, marbling, hardness, dryness, juiciness and pastiness; odor intensity; salty, sweet and bitter taste; flavor intensity and persistence; cured and rancid flavor. We used the squared correlation (R^2), to evaluate the quality of the prediction, calculated for each MRI slice (image) and for each meat piece. There is no single combination of texture feature vector and regressor which provides the best R^2 for the 39 attributes tested. On the contrary, the best combination is different for each attribute. However, the R^2 at meat piece level is from good to excellent ($R^2 > 0.56$, which correspond to $R > 0.75$) for 29 out to 39 attributes, from moderate to good for 8 attributes ($0.25 < R^2 < 0.56$) and only bad to moderate ($0.02 < R^2 < 0.25$) for the remaining 2 attributes (juiciness and flavor intensity in dataset H3). In relation with texture descriptors, the Gabor features (GNI vector) are the best for 5 attributes, the wavelet features (daub4, dtcwtVH, dtcwt and haar vectors) in 11 attributes, the fractal features (mfp and mfs vectors) in 7 atributtes, local binary patterns (lbp, mlbp, mlc and mlbc vectors) in 8 attributes and second-order statistics (vector coms, mcoms, sdhc, msdhc) in the remaining 8 attributes. In

relation to regressors, the multivariate adaptive regression splines (earth) and its Bagging ensemble (bagEarth) achieve the best result for 6 attributes; regularized least square regression (krlsRadial) is the best for 5 attributes; the support vector regression (svr), Bayesian regularized neural network (brnn), random forest (rf and RRF) and Bayesian additive regression tree (bartMachine), for 3 attributes each one.

Considering other measurements to evaluate the prediction quality, the WAPE (Weighted Absolute Percent Error) is lower than 2%, which is widely considered as a threshold of reliability, for 32 out of 37 attributes, excepting redness, marbling, hardness and rancid flavor of LSA, and rancid flavor of H3. Besides, difference between computer predictions and true values is not statistically significant for 28 out of 37 attributes according to a Wilcoxon signed ranksum test, but for 6 of the remaining 9 attributes with significant difference the dispersion of the computer predictions around the true values is lower than the variability of the true measurements over trails or expert's scores for physico-chemical or sensorial attributes, respectively. At the same time, the prediction of the majority of physico-chemical and sensorial attributes can be done using only one or three slices of the MRI sequence of the meat piece, instead of all slices of the sequence, which implies a strong reduction time of the scanning process. After integrating the proposed procedure in a software for the MRI computers, the desired analysis of each meat product could be carried out in real time (no more than 15 min. using 3 images), automatically and non-destructively, which would suppose an important saving in terms of time and money (one shorter analysis to determine lots of characteristics and the analysed meat products can be sold). This fact is very important for the future implementation of an automatic quality assessment system in meat industries.

This analysis leads to the conclusion that the combination of MRI images of meat pieces, image texture features and regression techniques can be an alternative technique to predict, with high reliability and in a non-destructive and innocuous way, the physico-chemical and sensorial attributes of meat. As well, it can be fast enough to operate as a quality assesment system in the meat industries. The future work will be to develop a software to automatically analyse MRI of loins and hams, which allows the on-line estimation of their quality characteristics in a non-destructive way. This software is intended to be included in the production systems of meat industries.

Acknowledments

The authors wish to acknowledge the funding received from the FEDER-MICCIN Infrastructure Research Project (UNEX-10-1E-402), Junta de Extremadura economic support for research group (GRU15173 and GRU15113), from the Xunta de Galicia (Centro singular de investigación de Galicia accreditation 2016-2019) and from the European Union (European Regional Development

Fund - ERDF). We also wish to thank the Animal Source Foodstuffs Innovation Service (SiPA, Cáceres, Spain) from the University of Extremadura and the Infanta Cristina University Hospital Radiology Service for the images acquisition of hams.

References

- [1] T. Pérez-Palacios, J. Ruiz, D. Martín, J. M. Barat, T. Antequera, Pre-cure freezing effect on physicochemical, texture and sensory characteristics of Iberian ham, *Food Sci. Technol. Int.* 17 (2) (2011) 127 – 133.
- [2] J. Ruiz, C. García, E. Muriel, A. I. Andrés, J. Ventanas, Influence of sensory characteristics on the acceptability of dry cured ham, *Meat Sci.* 61 (4) (2002) 347–354.
- [3] F. Toldrá, M. Flores, Y. Sanz, Dry-cured ham flavour: enzymatic generation and process influence, *Food Chem.* 59 (4) (1997) 523–530.
- [4] A. V. A. Resurrección, Sensory aspects of consumer choices for meat and meat products, *Meat Sci.* 66 (1) (2003) 11–20.
- [5] G. Gandemer, Lipids in muscles and adipose tissues changes during processing and sensory properties of meat products, *Meat Sci.* 62 (3) (2002) 309–321.
- [6] C. García, A. Carrapiso, *Tecnología del jamón Ibérico*, Mundi Prensa, Madrid, 2001, Ch. La calidad sensorial del jamón Ibérico y su evaluación: la cata y la cata del jamón, pp. 391–418.
- [7] T. Antequera, A. Caro, G. Rodríguez, T. Pérez-Palacios, Monitoring the ripening process of Iberian ham by computer vision on magnetic resonance imaging, *Meat Sci.* 76 (2007) 561–567.
- [8] P. Fantazzini, M. Gombia, P. Schembri, N. Simoncini, R. Virgili, Use of Magnetic Resonance Imaging for monitoring Parma dry-cured ham processing, *Meat Sci.* 82 (2) (2009) 219 – 227.
- [9] L. Manzocco, M. Anese, S. Marzona, N. Innocente, C. Lagazio, M. C. Nicoli, Monitoring dry-curing of S. Daniele ham by magnetic resonance imaging, *Food Chem.* 141 (3) (2013) 2246 – 2252.
- [10] T. Pérez-Palacios, T. Antequera, M. Durán, A. Caro, P. Rodríguez, J. Ruiz, MRI-based analysis, lipid composition and sensory traits for studying Iberian dry-cured hams from pigs fed with different diets, *Food Res. Int.* 43 (1) (2010) 248 – 254.
- [11] T. Pérez-Palacios, T. Antequera, M. Durán, A. Caro, P. Rodríguez, R. Palacios, MRI-based analysis of feeding background effect on fresh Iberian ham, *Food Chem.* 126 (3) (2011) 1366 – 1372.

- [12] E. Cernadas, P. Carrión, P. G. Rodríguez, E. Muriel, T. Antequera, Analyzing magnetic resonance images of Iberian pork loin to predict its sensorial characteristics, *Comput. Vis. Image Und.* 98 (2005) 345–361.
- [13] D. Caballero, T. Antequera, A. Caro, M. Ávila, P. G. Rodríguez, T. Pérez-Palacios, Non-destructive analysis of sensory traits of dry-cured loins by MRI-computer vision techniques and data mining, *J Sci. Food Agr.* 97 (9) (2017) 2942–2952.
- [14] D. Caballero, T. Pérez-Palacios, A. Caro, J. M. Amigo, A. B. Dahl, B. K. Ersbll, T. Antequera, Prediction of pork quality parameters by applying fractals and data mining on MRI, *Food Res. Int.* 99 (1) (2017) 739–747.
- [15] D. Caballero, A. Caro, P. G. Rodríguez, M. L. Durán, M. Ávila, R. Palacios, T. Antequera, T. Pérez-Palacios, Modeling salt diffusion in Iberian ham by applying MRI and data mining, *J Food Eng.* 189 (2016) 115 – 122.
- [16] T. Pérez-Palacios, D. Caballero, T. Antequera, M. L. Durán, M. Ávila, A. Caro, Optimization of MRI acquisition and texture analysis to predict physico-chemical parameters of loins by data mining, *Food and Bioprocess Technology* 10 (4) (2017) 750–758.
- [17] F. Bajd, M. Škrlep, M. Čandek Potokar, I. Serša, MRI-aided texture analyses of compressed meat products, *J Food Eng.* 207 (2017) 108 – 118.
- [18] T. Pérez-Palacios, D. Caballero, A. Caro, P. G. Rodríguez, T. Antequera, Applying data mining and Computer Vision Techniques to MRI to estimate quality traits in Iberian hams, *J Food Eng.* 131 (2014) 82 – 88.
- [19] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Intl. J. Conf. on Artif. Intel. (IJCAI)*, Montreal (Canadá), 1995, pp. 1137–1143.
- [20] T. Pérez-Palacios, J. Ruiz, D. Martín, E. Muriel, T. Antequera, Comparison of different methods for total lipid quantification in meat and meat products, *Food Chem.* 110 (4) (2008) 1025 – 1029.
- [21] N. Otsu, A threshold selection method from gray-level histograms, *IEEE T Syst. Man Cyb.* 9 (1).
- [22] A. Caro, P. Rodríguez, M. Durán, T. Antequera, Active Contours for Real Time Applications, in: M. Flournier (Ed.), *Perspectives on Pattern Recognition*, Nova Science Publishers, Inc., 2012, pp. 173–186.
- [23] E. Cernadas, M. Fernández-Delgado, E. González-Rufino, P. Carrión, Influence of normalization and color space to color texture classification, *Pattern Recogn.* 61 (2017) 120 – 138.

- [24] R. M. Haralick, K. Shanmugan, I. Dinstein, Textural features for image classification, *IEEE T Man Cyb.* 3 (6) (1973) 610 – 621.
- [25] M. Sonka, V. Hlavac, R. Boyle, Image processing, analysis, and machine vision, Int. Thomsom Publishing (ITP), 1999.
- [26] M. Unser, Sum and Difference Histograms for Texture Classification, *IEEE Trans. Pattern Anal. Mach. Intell.* 8 (1) (1986) 118–125.
- [27] T. Ojala, M. Piatikäinen, T. Mäenpää, Multiresolution grey-scale and rotation invariant texture classification with Local Binary Pattern, *IEEE T Pattern Anal.* 24 (7) (2002) 971–987.
- [28] C. Zhu, C.-E. Bichot, L. Chen, Image region description using orthogonal combination of local binary patterns enhanced with color information, *Pattern Recogn.* 46 (2013) 1949–1963.
- [29] Matlab, version 7.12 (R2011a), Natick (MA) (2011).
URL <https://es.mathworks.com/help/nnet>
- [30] Y. Zhao, D.-S. Huang, W. Jia, Complete local binary count for rotation invariant texture classification, *IEEE T Image Process.* 21 (10) (2012) 4492–4497.
- [31] M. Ivanovici, N. Richard, Fractal dimension of color fractal images, *IEEE T Image Process.* 20 (1) (2011) 227–235.
- [32] S. Jansson, Evaluation of methods for estimating fractal properties of intensity images, Ph.D. thesis, Umea University, Umea, Sweden (2006).
- [33] E. González-Rufino, P. Carrión, E. Cernadas, M. Fernández-Delgado, R. Domínguez-Petit, Exhaustive comparison of colour texture features and classification methods to discriminate cells categories in histological images of fish ovary, *Pattern Recogn.* 46 (2013) 2391–2407.
- [34] N. Sarkar, B. B. Chaudhuri, An efficient approach to estimate fractal dimension of textural images, *Pattern Recogn.* 25 (9) (1992) 1035 – 1041.
- [35] Y. Xu, H. Ji, C. Fermuller, Viewpoint invariant texture description using fractal analysis, *Int. J Comput. Vision* 83 (2009) 85–100.
- [36] J. S. Walker, A Primer on wavelets and Their Scientific Applications, Chapman Hall, 2008.
- [37] A. Laine, J. Fan, Texture classification by Wavelet Packet signatures, *IEEE T Pattern Anal.* 15 (11) (1993) 1186–1191.

- [38] I. W. Selesnick, R. G. Baraniuk, N. C. Kingsbury, The dual-tree complex wavelet transform, *IEEE Signal Proc. Mag.* 22 (6) (2005) 123–151.
- [39] T. Celik, T. Tjahjadi, T. Tjahjadi, Multiscale texture classification using dual-tree complex wavelet transform, *Pattern Recogn. Lett.* (2009) 331–339.
- [40] T. Randen, J. H. Husoy, Filtering for Texture Classification: A Comparative Study, *IEEE T Pattern Anal.* 21 (4) (1999) 291–310.
- [41] F. Bianconi, A. Fernández, Evaluation of the effects of Gabor filter parameters on texture classification, *Pattern Recogn.* 40 (2007) 3325–3335.
- [42] M. S. Sirsat, Application of machine learning to agricultural soil data, Ph.D. thesis, Computer Science School, CiTIUS (June 2017).
URL http://citius.usc.es/sites/default/files/tesis/Tese_ManishaSirsat.pdf
- [43] M. Sirsat, M. Fernández-Delgado, E. Cernadas, M. Febrero-Bande, S. Barro, An extensive experimental survey on regression techniques, Submitted.
- [44] M. Fernández-Delgado, M. Sirsat, E. Cernadas, S. Alawadi, S. Barro, M. Febrero-Bande, An extensive experimental survey of regression methods, *Neural Networks*.
URL <http://www.sciencedirect.com/science/article/pii/S0893608018303411>
- [45] R Core Team, R: A language and environment for statistical computing, Vienna, Austria (2008).
URL <https://www.R-project.org>
- [46] M. Kuhn, Caret: Classification and regression training, R package version 6.0-70 (2016).
URL <https://CRAN.R-project.org/package=caret>
- [47] D. Bates, J. Chambers, *Statistical models in S*, Wadsworth & Brooks/Cole, Pacific Grove, CA, 1992.
- [48] J. Goeman, L-1 penalized estimation in the Cox proportional hazards model, *Biometrical J.* 52 (2010) 70–84.
- [49] J. Hainmueller, C. Hazlett, Kernel regularized least squares: reducing misspecification bias with a flexible and interpretable machine learning approach, *Polit. Anal.* 22 (2013) 143–168.
- [50] T. Zhang, Adaptive forward-backward greedy algorithm for learning sparse representations, *IEEE T Inform. Theor.* 57 (7) (2011) 4689–4708.
- [51] D. Specht, A general regression neural network, *IEEE T Neural Netw.* 2 (1991) 568–576.

- [52] G.-B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, *IEEE Trans. Syst., Man, and Cybern.-Part B: Cybern.* 42(2) (2012) 513–529.
- [53] G. E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.* 18 (7) (2006) 1527–1554.
- [54] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, F. Alsaadi, A survey of deep neural network architectures and their applications, *Neurocomputing* 234 (2017) 11–26.
- [55] F. Chollet, Keras, <https://keras.io> (2015).
- [56] C. Chang, C. Lin, LIBSVM: a library for support vector machines, *ACM T Intel. Syst. Tec.* 2 (2011) 27:1–27:27.
- [57] R. Quinlan, Learning with continuous classes, in: *Proc. Australian J. Conf. on Artif. Intel.*, 1992, pp. 343–348.
- [58] R. Quinlan, Combining instance-based and model-based learning, in: *Proc. Intl. Conf. on Mach. Learn.*, 1993, pp. 236–243.
- [59] J. Friedman, Multivariate adaptive regression splines, *Ann. Stat.* 19 (1) (1991) 1–141.
- [60] P. Buehlmann, B. Yu, Boosting with the L2 loss: regression and classification, *J. Am. Stat. Assoc.* 98 (2003) 324–339.
- [61] L. Breiman, Random forests, *Machine Learning* 45 (1) (2001) 5–32.
- [62] M. Kursa, W. Rudnicki, Feature selection with the Boruta package, *J. Stat. Softw.* 36 (11) (2010) 1–13.
- [63] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, *Mach. Learn.* 63 (1) (2006) 3–42.
- [64] I. Mizera, R. Koenker, Convex optimization in R, *J. Stat. Softw.* 60 (5) (2014) 1–23.
- [65] F. Foresee, M. T. Hagan, Gauss-Newton approximation to Bayesian regularization, in: *Intl. J. Conf. on Neural Netw.*, 1997, pp. 1930–1935.
- [66] D. MacKay, Bayesian interpolation, *Neural Comput.* 4 (1992) 415–447.
- [67] A. Kapelner, J. Bleich, bartMachine: machine learning with Bayesian additive regression trees, *J. Stat. Softw.* 70 (4) (2016) 1–40.
- [68] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Stat.* 32 (2004) 407–499.

- [69] J. Friedman, W. Stuetzle, Projection pursuit regression, *J. Am. Stat. Assoc.* 76 (1981) 817–823.
- [70] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, *J. R. Stat. Soc.* 67 (2005) 301–320.
- [71] J. D. Gibbons, S. Chakraborti, *Nonparametric Statistical Inference*, CRC Press, 2011.
- [72] T. Colton, *Statistical in medicine*, Little Brown and Co., New York, NJ, 1974.
- [73] S. García, A. Fernández, A. Benítez, F. Herrera, Statistical comparisons by means of non-parametric tests: a case study on genetic based machine learning, in: *Proc. of the II Congreso Español de Informática (CEDI 2007)*, 2007, pp. 95–104.