HYPERSPECTRAL IMAGING FOR ASSESSING THE QUALITY ATTRIBUTES OF CURED PORK LOIN

Andreas Kartakoullis, Andreas Kamilaris, Joel Gonzalez, Xavier Serra, Pere Gou, Maria Font

Department of Food Technology, IRTA, Monells, Spain
andreas.kartakoullis@irta.cat

ABSTRACT

Meat is one of the most widely consumed products in the world and its market value is strongly correlated with its quality properties such as chemical composition, technological and sensory attributes. Therefore, food industry needs to minimize the variability of meat quality properties in order to keep the quality standards as high as possible. The current industrial methods for assessing the meat quality properties are destructive, time consuming and consequently unsuitable for the on-line application classification of meat quality properties. On the other hand Hyperspectral Imaging (HSI) allows on-line high-throughput screening in a non-destructive nature. We have use HSI with random forest classification on cured pork loin meat for classify their quality standards with a classification accuracy of 95%.

Index Terms— Visible Hyperspectral Imaging, Random Forest Classification, Food Technology

1. INTRODUCTION

A hyperspectral image consists of a high number of spectrally well-separated narrow bands. The whole hyperspectral image, consisting of all these spectral bands of each image, is also called the image hypercube, a three dimensional dataset having two axes of spatial information and one axis of spectral information. Each hyperspectral pixel for reflection measurements, represents the reflection spectrum of the corresponding spot in the region of interest. Because of this high-throughput ability to combine imaging with spectroscopy HSI has become more and more popular in food quality control in order for industry to meet consumer demands and the challenge of market segmentation and legal restrictions [1].

The quality of meat depends on a great number of factors such as farm practices, animal background and handling from farm to slaughter. The most widely used parameters in pork quality are pH, water holding capacity (WHC), color and firmness. The standards to determined these parameters can be different and usually is a combination of different chemical and sensorial measurements according to need of the market [2]. For industrial characterisation three main classes are described in the literature as follows: PSE (pale, soft, exudative), PFN (Pale, firm, non-exudative), RSE (red, soft, exudative), RFN (red, firm, non-exudative) and DFD (dark, firm, dry), resulting from the combination of pH, WHC and/or lightness of the colour ($L^*$ value) [3].

In the current study, we propose a novel supervised machine learning method to classify RFN, PFN and PSE cured pork loins from hyperspectral images in the visible region. The cured pork loins were acquired and characterized from meat companies in the area of Catalonia, Spain. The random forest classification have resulted into a classification accuracy of 95%, which is encouraging enough for pursuing a bigger project at an industrial scale.

2. DATA AND METHODS

In total sixteen different whole pork loins were acquired from three different meat companies that were characterized accordingly in the three different categories of RFN, RSE and PFN. The loin were dissected and set for image acquisition by removing any traces of humidity from their surface. The hyperspectral images of this study were acquired by a Resonon Pika-L Vis-Nir hyperspectral imaging camera operating at the range of 400 – 1000 nm with a spectral resolution of 2.1nm, 300 spectral channels and 900 spatial channels. The hyperspectral images were acquired in a line mapping configuration with the camera placed in a right angle respect to the sample. The original dimensions of each image was 800x900 pixels and the original hypercube had dimensions of 800x900x300 and size of 500Mbs. Spatial and spectral binning was performed on each image in order to reduce their dimensions and size down to 85x95x30 and 7.2Mbs. The preprocess of the hyperspectral images was done by the Resonon SpectrononPro software and the analysis by in-house developed codes under R programming language using the randomForest package.

Hyperspectral images are high-dimensional data containing many hidden information and patterns, and influenced by many external factors. The following steps were taken for the analysis of the hyperspectral images: (i) data pre-processing
to eliminate noise instrument, background and highlight the spectral features for the next step (ii) train predictive models using a set of samples with known characteristics obtained by suitable reference methods, and (ii) cross-validate the predictive models using part of the images that was kept outside of the training set.

For the first step the removal of dead pixels and spikes was done together with spatial and spectral binning and subsequently was the background removal and the spatial pre-processing. This part is to select the Region of Interest (ROI) in the image by masking the non-interesting areas. K-means clustering was first applied by using 2 clusters to remove the background and keeping only the cured loins as Figure 1 depicts. The main feature that we focused to classify the images was the type and amount of pigment that the original meat had before the cured process. We selected myoglobin as the main feature because has different levels and type in the three quality categories that we are interested.

Thereafter, a second K-means clustering with 3 clusters was performed to the first-derivative of the spectra of each loin to determine the ROI. The center of each loin was selected to the fact that myoglobin has a higher concentration there. In addition to the k-means clustering, a morphological close operation was applied, which is a dilation followed by an erosion to fill gaps in the images. Figure 2 shows the selected ROIs of the loins after the whole procedure was completed.

The second step was to build the predictive model by first partitioning the spectra (80% − 20%) into training and test-
Instead of random sample section we have used the Kennard-Stone, which is a sequential method that covers the spectral space uniformly in order to build a training dataset with flat distribution over the spectral space. It starts by selecting two spectra that are the farthest apart from each other and the next spectra is the one that is the most distant from the previous ones [5]. We have used the Mahalanobis geometric distance for selecting the which spectra will go for training and which for testing.

Next was the spectral pretreatment of the images in order to highlight the spectral features that will be used to train the predictive model. We applied Savitzky-Golay filters to obtain the first derivative of each spectra. The derivatives change the shape of the spectra, making more difficult their analysis. Therefore, the selection of the proper window for fitting the polynomial plays a fundamental role. Choosing small windows may generate noise in the spectra; whereas the selection of a wide window may eliminate important information if the spectra contain sharp peaks [6]. Through the cross-validation scheme of the third step we have selected a Savitzky-Golay filter with a window size of 5, second order polynomial degree and first order derivative degree.

The third step for building the predictive model started with the data dimensionality reduction with the use of Principal Component Analysis (PCA). Figure 3 shows the scatter plot of the PC1 and PC2 scores and the three classes of the meat quality have been differentiated quite well. The PCA model was built with the first derivative and mean centered spectra. Figure 3 shows two subclasses within the PFN class (red colour) that is attributed to the different genetic origin of the animals that were used for the production of the loins. Thereafter, was the application of the random forest multiclass classifier. Random forest methods are supervised learning algorithms that are an ensemble of decision trees, meaning that the random forest algorithm builds several decision trees and merge them together to get a more accurate and stable prediction [7]. It consider to be one of the most flexible and easy use machine learning algorithm due to the fact that does not need hyper parameter tuning. Additionally, it adds randomness to the model while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model.

### 3. RESULTS

For building the model we used the dimensionality reduced data from the PCA process as an input to random forest classifier with 250 trees to grow and with 3 number of variables randomly sampled as candidates at each split. The different PCs scores of the training dataset were added each time to the classifier until to find the minimum of the out-of-bag error. The optimum model was found to be with the first 5 PCs corresponding to more than 90% of the variance, and with 200 trees to grow and 4 randomly sampled variables. Table 1 shows the confusion matrix of the random forest classification on the testing dataset that was performed on the PCA-reduced data with classification accuracy of 95.2% and with Cohen’s kappa \( \kappa = 0.92 \). The Cohen’s kappa statistic is a very good metric for multiclass classification problems that accuracy and F-measure can not provide a complete picture of the classification performance especially if we have a classification problem with unbalanced classes. Cohen’s kappa is defined as:

\[
\kappa = \frac{p_o - p_e}{1 - p_e}
\]

where \( p_o \) is the observed agreement and \( p_e \) is the expected agreement. Its value can be interpret as follows: (i) < 0 is indicating no agreement, (ii) 0 – 0.20 slight agreement, (iii) 0.21 – 0.40 fair agreement, (iv) 0.41 – 0.60 moderate, (v) 0.61 – 0.80 substantial, (vi) 0.81 – 1.0 almost perfect agreement [8]. In our case the high value of Cohen’s kappa indicates almost perfect agreement.

### 4. CONCLUSIONS

In this study, we have demonstrated that the meat quality attributes of pork loins can be classified with high accuracy by using hyperspectral imaging in the visible part of the spectrum. The results are encouraging enough to launch a bigger project that will consider also the genetic origin of the animal as well their breeding practises.

### 5. ACKNOWLEDGEMENTS

Andreas Kartakoullis and Andreas Kamilaris acknowledge financial support from the Marie Sklodowska-Curie COFUND P-SPHERE project, which has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 665919.
6. REFERENCES


[5] Tomasz Puzyn; Aleksandra Mostrag-Szlichtyng; Agnieszka Gajewicz; Michal Skrzynski; Andrew P. Worth, “Investigating the influence of data splitting on the predictive ability of qsar/qspr models,” *Structural Chemistry*, vol. 22, 08 2011.

