Non-Destructive Prediction of Pork Meat Degradation using a Stacked Autoencoder Classifier on Hyperspectral Images

B. Gallo, Sergio de Almeida, José-Carlos M Bermudez, Chen Jie, Cédric Richard

Non-Destructive Prediction of Pork Meat Degradation using a Stacked Autoencoder Classifier on Hyperspectral Images

B. B. Gallo
Catholic University of Pelotas
Pelotas, Brazil
bettybraga@yahoo.com.br

S. J. M. de Almeida
Catholic University of Pelotas
Pelotas, Brazil
sergio.almeida@ucpel.edu.br

J. C. M. Bermudez
Federal University of Santa Catarina
City, Country
j.bermudez@ieee.org

J. Chen
Northwestern Polytechnical University
Xi'an, China
dr.jie.chen@ieee.org

C. Richard
Université de Nice Sophia-Antipolis
Nice, France
cedric.richard@unice.fr

***Abstract*—This work presents initial results on a multitemporal hyperspectral image analysis method to evaluate the time degradation of pork meat. The proposed method is inexpensive and practically non-destructive. The hyperspectral data is analyzed and the relevant information is reduced to the information in only three wavelengths. The analysis is performed by a binary classifier composed by two stacked autoencoders and a softmax output layer. The use of autoencoders reduces tenfold the dimension of the input space. The proposed classifier has led to 97.2% of correct decisions, which indicates the great potential of the methodology to monitor the safety of meat.**

***Index Terms*—Hyperspectral imaging, meat quality assessment, machine learning, neural network.**

I. Introduction

Non-destructive and fast methods to predict meat quality and safety attributes have recently become a major research objective, as classical analysis methods are quite invasive and destroy part of the meat [1]. Consumer market safety and quality demands require that meat quality be assessed in all industrial processes. High health risks accrue whenever safety requirements are not properly observed. According to a World Health Organization (WHO) report (2015), about 600 million (approximately one in ten people in the world) become ill after consuming contaminated food and 420,000 die each year [2], [3]. Among the several methods recently proposed to evaluate meat quality, the non-destructive analyses of scanned images have been the most promising ones. The use of hyperspectral images combines conventional digital imaging and spectroscopy. It consists in acquiring the spectral characteristics of a material at different wavelengths. The main difference between the various types of images (e.g. panchromatic, multispectral and hyperspectral) is the number of spectral bands acquired. In hyperspectral analysis, information on a particular material is acquired by means of electromagnetic radiation sensors, usually on the portion of the spectrum that extends from the visible to the infrared region. The basic principle is the fact that all materials reflect electromagnetic energy, at specific wavelengths and in distinct patterns depending on their molecular structure. The results obtained are in the form of a three-dimensional cube, with two spatial and one spectral dimensions, and each pixel of the recorded image corresponds to a reflectance spectrum. Most hyperspectral imaging algorithms aim to detect objects or materials, known or unknown, in a given scenario, classify or segment the image in regions where certain types of material predominate, or estimate the distribution of the quantities of different materials in a pixel of the image [4], [5], [6] and [7]. Hyperspectral unmixing is the procedure by which the measured spectrum of a mixed pixel is decomposed into its constituent members, known as endmembers, and the fraction of each endmember in a pixel (known as abundance) is estimated [5].

Hyperspectral images have been recently applied to food quality assessment [10], [11], and more specifically to meat quality assessment [12] [3]. An overview of wavelength selection techniques for hyperspectral image processing in the food industry can be found in [13]. One important aspect of meat quality assessment is the prediction of meat degradation in time. For instance, [14] studied the water distribution in beef during dehydration using time series of hyperspectral images.

This paper presents initial results on non-destructive prediction of pork meat degradation from hyperspectral images. After a careful selection of the most representative spectral bands for determining meat degradation, we propose a classifier using a stacked autoencoder to reduce the dimension of the data space. Hence, the supervised training of the classifer is performed only on the parameters of the output softmax layer. Simulation results illustrate the high classification accuracy that can be obtained using the proposed solution.

This work was partly supported by CNPq under grants Nos. 304250/2017-1 and 409044/2018-0.

The remainder of this paper is organized as follows. Section II describes the data acquisition process. Section III presents the spectral analysis and the selection of relevant spectral information. Section IV and Section V detail the network structure and training, as well as the performance obtained. Section VI concludes and discusses future work.

II. Materials and Methods

Seven hyperspectral images of samples of pork meat were acquired in a 24-hour period, with an ambient temperature of 26 degrees Celsius. The data were collected by the GaiaField and GaiaSorter systems at NPU. Our GaiaField (Sichuan Dualix Spectral ImageTechnology Co. Ltd., GaiaField-V10) is a push-boom imaging spectrometer with a HSIA-OL50 lens, covering the visible and NIR wavelengths ranging approximately from 390nm to 1000nm, with a spectral resolution of up to 0.58nm. GaiaSorter sets an environment that isolates external lights, and is endowed with a conveyor to move samples for the push-boom imaging. Four tungsten-bromine lamps were used to form a hemispherical-directional illumination.

We applied the black-white normalization to the raw data to convert the collected light intensity levels to reflectance values, and to remove the effect of the dark current of the camera sensor and avoid the uneven light intensity of each band. In the offline phase, the black image is acquired by turning off the light source and covering the camera lens with its cap. The white image was acquired by imaging a standard white board (foamed PEFE resin, certified by China Metrology Institute) under the same condition as that of the raw image for calibration.

Fig. 1 shows a time line with the moments of acquisition of each image. The samples were named as M1, M2, M3, M4, M5, M6 and M7. Note that the time interval between acquisitions varies between 6 and 12 hours. Upon careful examination of the meat samples it was possible to detect changes in the appearance and odor of the meat samples starting at sample M4, which identified the beginning of the deterioration process. Hence, it was considered that the meat should be classified as unfit for consumption after the first 24 hours.

Each image contained nine separate pieces of pork, as shown in Fig. 2. We selected three pieces of meat with low fat concentration for analysis. The same locations were considered for the seven time samples M1-M7. The corresponding areas are indicated by arrows in Fig. 2.

These pieces were initially analyzed to determine the wavelengths that provided more information for the purpose of classification. In this step, we defined one hypercube with dimension $10\times10\times256$ pixels in each of the three pieces, and in each of the seven hyperspectral images. The complete

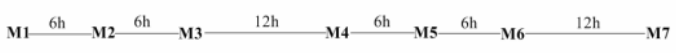

Fig. 1. Time interval between acquired meat images.

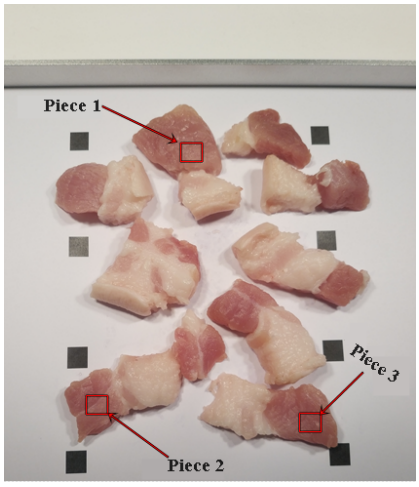

Fig. 2. RGB image of one time sample. Little squares indicate the three pieces selected for analysis. The same locations were used in all time samples. The upper part of the image is the standard white board used for data normalization.

image has 696×790 pixels. A classic endmember extraction algorithm was then applied to each hypercube to reduce the effect of noise in the samples, and to explore the existence of different component spectra. After determining the best set of wavelengths, we trained the classifier using reduced hypercubes, as explained in the following.

III. Spectral Analysis

The endmember extraction was performed using the Vertex Component Analysis (VCA) algorithm [8]. The extraction using the 390nm to 1000nm wavelength range resulted in a set of endmembers that evidently represented a single endmember and with spatial variability. We have then defined a single average endmember for each piece (1-3) of a given time sample (M1-M7). Figs. 3 to 5 show the spectral signatures of the endmembers of the temporal sequence of pieces 1-3 indicated in Fig. 2.

In Fig. 3 we observe that the reflectances in the visible range from 450nm to 750nm have larger values for the M1-M3 images. Also, the reflectance values decrease with time from M1 to M7. We also note that the endmembers for the

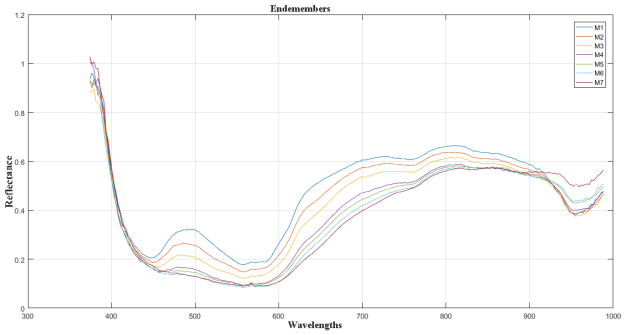

Fig. 3. Average endmembers extracted from piece 1 for each of the seven images of the sequence M1 to M7.

M4-M7 range (deteriorated meat) are more clustered in a wide spectral range. These characteristics are strongly modified in the wavelength range from 920nm to 990nm (near infrared - NIR). In Figs. 4 (piece 2) and 5 (piece 3) the reflectance behaviors for the set M1-M3 and for the set M4-M7 are clearly distinct for the whole spectrum, with M4 usually in the boundary between the two groups. From these curves and observations we chose to select three wavelengths for which the reflectances presented better regularity in behavior both in the visible range of the spectrum and in the NIR region. The selected wavelengths were 500nm, 700nm and 960nm. Note that the first two are in the visible range and the third in the NIR range.

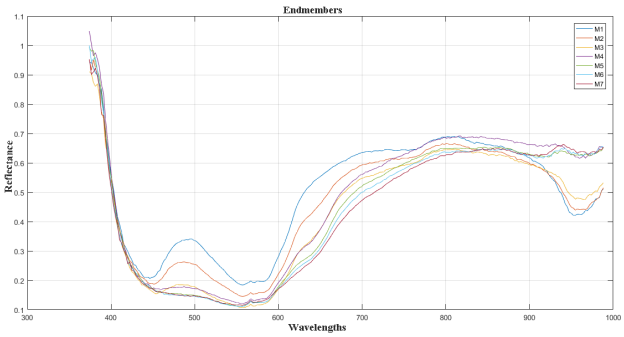

Fig. 4. Average endmembers extracted from piece 2 for each of the seven images of the sequence M1 to M7.

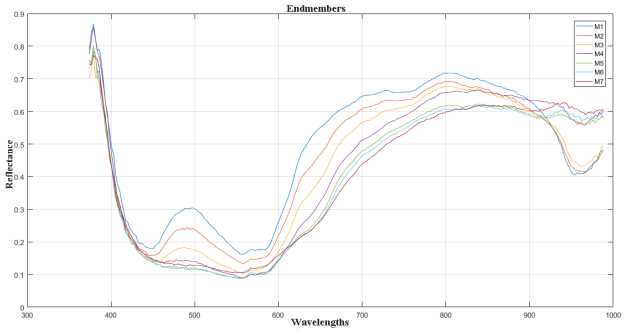

Fig. 5. Average endmembers extracted from piece 3 for each of the seven images of the sequence M1 to M7.

IV. Neural Network Structure

Having selected the three wavelengths to be used for classification purposes, we next defined the structure of the neural network (NN) classifier as follows:

a) The inputs to the NN were 100 $10 \times 10 \times 3$ hypercubes, each arranged as a 300×1 input vector, with one 100×1 sub-vector for each wavelength.
b) There were a total of 306 hipercubes, 135 from edible meat (M1-M3) and 171 from degraded meat (M4-M7).
c) 60% of the data (183 samples) were used for training and 40% (123 samples) were used for test.
d) To reduce the dimension of the data space, we built the neural network with two stacked autoencoders, followed by a softmax layer to perform a binary classification.
e) The two output classes were "edible meat" and "degraded meat".

A. The Autoencoders

Autoencoders are generative models composed of an encoder and a decoder [16]. Fig. 6 illustrates the principle. The encoder has the purpose of mapping the inputs x to a hidden representation y, usually of lower dimension, whereas the decoder tries to reverse this mapping, with the objective of reconstructing the original input in z. Usual cost functions are the mean square error (MSE) and the cross entropy. The training is unsupervised. .

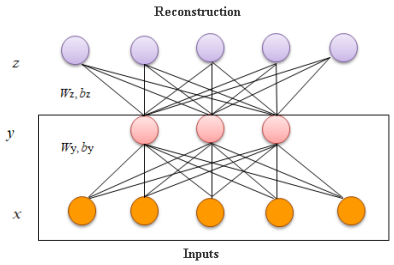

Fig. 6. Reconstruction of the input.

A stacked autoencoder consists of multiple layers of sparse autoencoders in which the outputs of each layer are wired to the inputs of the successive layer. We have used two stacked autoencoders, trained using the greedy layer-wise approach. We first trained the first layer on the original inputs (300×1 vectors) to obtain 100×1 reduced dimension representations[1]. Then, we removed the decoder of the first autoencoder, freezed the encoder parameters, and used its 100×1 outputs as inputs to train the second autoencoder, yielding 30×1 reduced dimension representations (again dimension empirically determined). Both trainings were performed using backpropagation. Then, the two trained encoders were connected in tandem to an output softmax layer. Fig. 7 illustrates the structure.

Both autoencoders were implemented using the sigmoid activation function

$$f(s) = \frac{1}{1 + e^{-s}} \tag{1}$$

where s is the input to the activation function $f(s)$. To determine the best cost function for training the autoencoders, we fixed the softmax cost function as the cross entropy, and performed the complete training using the MSE and the cross entropy as cost functions for the autoencoder training (details of the test procedure below). After several tests, the MSE cost function was adopted since it has led to the best average performance results. All the training data have been used to train the autoencoders (no validation or test sets). Table I

[1]This dimension has been defined after tests with several dimensions.

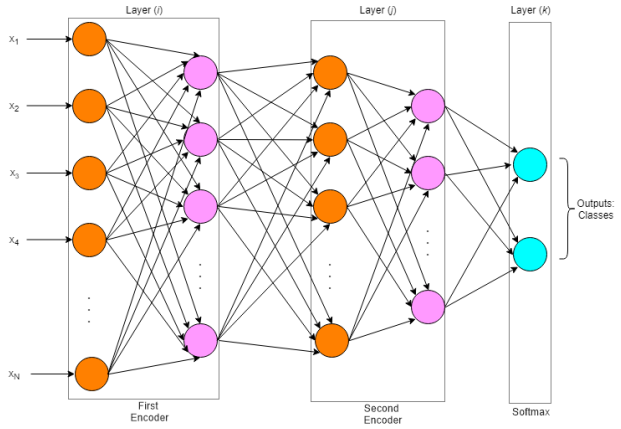

Fig. 7. Neural Network Structure.

shows the obtained percentages of correct classifications using the two cost functions to train the autoencoders.

TABLE I
COMPARISON OF COST FUNCTIONS TO TRAIN THE AUTOENCODERS.

Cross Entropy	85.87%
Mean Square Error	97.20%

B. The output layer

The softmax layer has been trained is a supervised manner using backpropagation after freezing the parameters of the two trained encoders. The cost function used was the cross-entropy function which, after the tests performed, showed a better classification performance than the MSE function. For these tests, the encoders trained with the MSE cost function were employed. No overall tunning was performed. Table II shows the percentages of correct classifications training the softmax layer (and thus the classifier) using the two cost functions.

TABLE II
COMPARISON OF COST FUNCTIONS TO TRAIN THE SOFTMAX LAYER.

Cross Entropy	97.20%
Mean Square Error	96.16%

V. NETWORK TRAINING

The network parameters have been initialized with random values to break the symmetry in training and avoid lockup [16]. Other important details of the training process are as follows:

a) The training was composed of iterations and epochs. Each iteration corresponds to the presentation of the complete training set, with the sequence of presentation of the input vector randomized before each iteration. For each input vector several epochs were realized to define the parameter values for that iteration.
b) For each iteration, a different training set with 183 samples and a test set with 123 samples were randomly chosen.
c) The number of iterations and epochs has been determined by cross-validation (explained later).
d) After each iteration, the network parameters were frozen, and the classification performance was verified using the test data (123 test vectors).
e) The classification performances reported are the averages of the performances obtained in all iterations.

A. Number of iterations and epochs

To set the number of iterations and epochs to be employed in the training of the network, we used cross-validation. First, using a very large number of epochs per iteration, we performed the training and the test of the classifier for several numbers of iterations. The classification results obtained are shown in Fig. 8. From these results, we decided to set the number of iterations to 500.

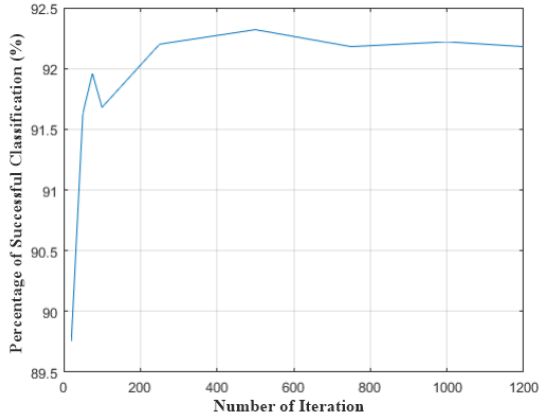

Fig. 8. Classification performance versus number of iterations.

With this number of iterations fixed, we have again performed training and test of the classifier for different numbers of epochs per iteration. The obtained classification results are shown in Fig. 9. Given these results we opted to perform training of the networks using 500 iterations and 125 epochs per iteration. The results reported in Table I were obtained for

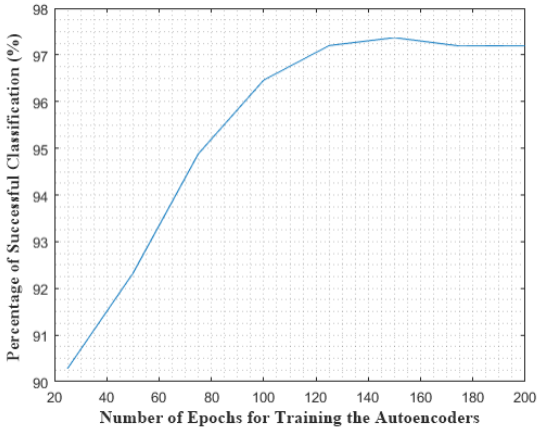

Fig. 9. Classification performance versus number of epochs for 500 iterations.

these values.

VI. Conclusions

This paper has presented a non-invasive method to predict the degradation of meat quality over time using hyperspectral images. In these initial results we have used image acquisition in the visible and near-infrared spectral regions, which can be obtained with less expensive cameras. Analysis of the hyperspectral images permitted the reduction of the relevant information to a set of only three wavelengths. The training supervision was based of information appearance and odor of the meat samples. The classifier was implemented using two stacked autoencoders to reduce ten times the dimension of the input space, and a softmax output layer. Training cost functions and parameters have been determined by cross-validation. The average performance of 97.2% of correct classification that has been obtained illustrates that the proposed method has a good potential for the intended application. More work is necessary to refine the data acquisition process, using a higher sampling rate, and to improve the determination of the target values using more technically supported methods. These are objectives of future work, given the initial success. Nevertheless, the proposed method allows for a very inexpensive first evaluation of meat quality over time, which can then be sophisticated as initial degradation evidence is detected.

References

[1] Tao, Feifei, and Yankun Peng. "A method for nondestructive prediction of pork meat quality and safety attributes by hyperspectral imaging technique." Journal of Food Engineering 126 (2014): 98-106.

[2] WHO 2015, World Health Organization, Global and regional food consumption patterns and trends. Available at: http://www.who.int/nutrition/topics/3foodconsumption/en/index4.html (2015).

[3] Kozan, Hasan and Sarıçoban, Cemalettin and AKYÜREK, Hasan and Ünver, Ahmet. "Hyperspectral imaging technique as a state-of-art technology in meat science." Green Chemistry and Technology Letters. (2016).

[4] Bioucas-Dias, José M., et al. "Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 5.2 (2012): 354-379.

[5] Borengasser, Marcus, William S. Hungate, and Russell Watkins. Hyperspectral remote sensing: principles and applications. CRC press, 2007.

[6] Keshava, Nirmal, and John F. Mustard. "Spectral unmixing." IEEE Signal Processing Magazine, 19.1 (2002): 44-57.

[7] Bioucas-Dias, José M., et al. "Hyperspectral remote sensing data analysis and future challenges." IEEE Geoscience and Remote Sensing Magazine, 1.2 (2013): 6-36.

[8] Nascimento, José M. P., and Bioucas-Dias, José M. "Vertex component analysis: a fast algorithm to unmix hyperspectral data." IEEE transactions on Geoscience and Remote Sensing, 43.4 (2005): 898-910.

[9] Lin, Zhouhan, et al. "Spectral-spatial classification of hyperspectral image using autoencoders." 2013 9th International Conference on Information, Communications and Signal Processing. IEEE, 2013.

[10] Wu, Di, and Da-Wen Sun. "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review — part II: applications." Innovative Food Science and Emerging Technologies 19 (2013): 15-28.

[11] Huang, Hui, Li Liu, and Michael Ngadi. "Recent developments in hyperspectral imaging for assessment of food quality and safety." Sensors 14.4 (2014): 7248-7276.

[12] Xiong, Zhenjie, et al. "Recent developments of hyperspectral imaging systems and their applications in detecting quality attributes of red meats: a review." Journal of Food Engineering, 132 (2014): 1-13.

[13] Liu, Dan, Da-Wen Sun, and Xin-An Zeng. "Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry." Food and Bioprocess Technology, 7.2 (2014): 307-323.

[14] Wu, Di and Wang, Songjing and Wang, Nanfei and Nie, Pengcheng and He, Yong and Sun, Da-Wen and Yao, Jiansong. "Application of time series hyperspectral imaging (TS-HSI) for determining water distribution within beef and spectral kinetic analysis during dehydration." Food Bioprocess Technol, (2012): July 19 (online).

[15] Corinna Cortes and Vladimir Vapnik. "Support-vector networks." Machine Learning, 20(3):273–297, Sep 1995.

[16] Andrew Ng. "Sparse autoencoder." In CS294A Lecture note, pages 1-19. Stanford University, 2011.