



Neural networks applied to discriminate botanical origin of honeys



Ofélia Anjos^{a,b,*}, Carla Iglesias^c, Fátima Peres^a, Javier Martínez^d, Ángela García^c, Javier Taboada^c

^a Instituto Politécnico de Castelo Branco, Escola Superior Agrária, Apartado 119, 6001-909 Castelo Branco, Portugal

^b Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisboa, Portugal

^c Department of Natural Resources and Environmental Eng., University of Vigo, 36310 Vigo, Spain

^d Centro Universitario de la Defensa, Academia General Militar, 50090 Zaragoza, Spain

ARTICLE INFO

Article history:

Received 12 August 2014

Received in revised form 19 November 2014

Accepted 20 November 2014

Available online 27 November 2014

Keywords:

Honey

Physical–chemical parameters

Botanical origin

Neural networks

Overfitting

Classification problem

ABSTRACT

The aim of this work is develop a tool based on neural networks to predict the botanical origin of honeys using physical and chemical parameters.

The managed database consists of 49 honey samples of 2 different classes: monofloral (almond, holm oak, sweet chestnut, eucalyptus, orange, rosemary, lavender, strawberry trees, thyme, heather, sunflower) and multifloral. The moisture content, electrical conductivity, water activity, ashes content, pH, free acidity, colorimetric coordinates in CIELAB space (L^* , a^* , b^*) and total phenols content of the honey samples were evaluated. Those properties were considered as input variables of the predictive model. The neural network is optimised through several tests with different numbers of neurons in the hidden layer and also with different input variables.

The reduced error rates (5%) allow us to conclude that the botanical origin of honey can be reliably and quickly known from the colorimetric information and the electrical conductivity of honey.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Honey is a natural food product processed by honeybees by blending the sweetened sap collected from flowers with metabolic gastric enzymes. It is a very rich food product due to its composition, which contains water, saccharides (mainly fructose and glucose), minerals, amino acids and proteins (Anjos, Campos, Ruiz, & Antunes, 2015; Doner, 1977; Mesallam & El-Shaarawy, 1987). The differences observed in the honey composition depend on the region, season, nectar plant source, bee's breed (Cavaco, Miguel, Antunes, & Guerra, 2012).

The main quality parameters for honey are given by Codex Alimentarius (2001) as follows: moisture content, acidity, hydroxymethylfurfural content (HMF), diastase activity and sugars content (sum of glucose, fructose and sucrose). The honey composition is related to its characteristics, namely, the profiles of predominant sugars (glucose, fructose, sucrose and maltose) that have been associated with a wide variety of second-order quality properties such as viscosity, hygroscopic, granulation and energy value (Ouchemoukh, Schweitzer, Bachir Bey, Djoudad-Kadji, &

Louaileche, 2010). Another important parameter regarding food, especially honey, is color, which is often related to the consumption patterns and consequently used to make judgments on its quality (Caivano & Buera, 2012; Wilczyńska, 2014).

Depending on the botanical origin of honey, two main types can be distinguished: monofloral or unifloral honey when there is a predominant plant species' nectar, and multifloral honey when the nectar is collected from more than one plant species. Unifloral honey types have higher market value due to their limited production and availability, and their geographical origin can also be an important economical factor due to the existence of protected designation of origin (PDO) and protected geographical indications (PGI) (Cajka, Hajslova, Pudil, & Riddellova, 2009). The determination of the geographical and botanical honey origin is regulated by specific legislation (Codex Alimentarius Commission FAO/OMS, 2001; European Union, 2002).

Melissopalynological analysis is the reference method to determine the botanical origin of honey. However, this method requires a specific background and very well trained technicians with a high expertise in the interpretation of results. It is also necessary to identify a large amount of pollen grain, which makes the process of the identification of botanical origin of honey relatively difficult and very time-consuming (Benedetti, Mannino, Sabatini, & Marcazzan, 2004). Due to this fact, there is a need for alternative methods to analyse the botanical origin of honey in a faster, easier

* Corresponding author at: Instituto Politécnico de Castelo Branco, Escola Superior Agrária, Apartado 119, 6001-909 Castelo Branco, Portugal. Tel.: +351 272339900; fax: +351 272339901.

E-mail address: ofelia@ipcb.pt (O. Anjos).

way for sample screening in a quality control laboratory (Corvucci, Nobili, Melucci, & Grillenzoni, 2015; Escriche, Kadar, Domenech, & Gil-Sánchez, 2012).

Some methods are being developed for this proposes: electronic devices such as the electronic-tongue (e-tongue) or the electronic-nose (e-nose) with chemometric models have been proved to be useful tools for the discrimination of honeys' botanical origin (Benedetti et al., 2004; Cavaco et al., 2012; Dias et al., 2008; Escriche et al., 2012; Ulloa et al., 2013).

In this work, we use neural networks to determine the botanical origin of honey samples. Previous research works have proved they are an effective method for managing natural parameters due to their ability to find the complex relationships between the studied variables (Cilla, Martínez, Peña, & Martínez, 2012; Iglesias et al., 2014). Regarding honey, neural networks have also been applied: Benedetti et al. (2004) and Escriche et al. (2012) used an electronic nose to obtain the volatiles profile and neural networks to discriminate the origin of the samples; Shafiee, Minaei, Moghaddam-Charkari, and Barzegar (2014) characterised honey through ash content, antioxidant activity and total phenolic content, which were estimated using neural networks and color information.

Other research works have focused on the study of color to determine the botanical origin since it is related to the chemical composition of the product (González-Miret, Terrab, Hernanz, Fernández-Recamales, & Heredia, 2005; González-Paramás, García-Villanova, Gómez Bárez, Sánchez Sánchez, & Ardanuy Albajar, 2007; Gámbaro, Ares, Giménez, & Pahor, 2007; Juszczak, Socha, Rożnowski, Fortuna, & Nalepka, 2009; Terrab, Diez, & Heredia, 2002; Tuberoso et al., 2014).

The aim of this work is to evaluate the performance of neural networks for the determination of the botanical origin of honey samples from some physical–chemical properties. For this purpose, the best input parameters are evaluated in order to achieve a successful classification with the lowest number of parameters. The objective is to propose an objective methodology to evaluate the botanical origin of honey based on the results of quality control analyses.

The importance of this study could be the application of this technique to a fast screening in a quality control laboratory in order to identify the botanical origin of honey. Given the specificity of the botanical analysis and their slowness, this technique could be helpful in this determination in the future, with great potential for the melissopalynology classification of honey samples.

2. Materials and methods

2.1. Honey samples

The monofloral honey samples produced in different Portuguese and Spanish regions were obtained from the commercial market. The samples were stored in the dark at a room temperature of 12 °C. A total of 49 samples were taken from 14 different botanical origins.

The preparation of honey samples for identification of botanical origin followed the standardized acetolyzed method of Erdtman (1960).

20 g of honey sample was dissolved in 40 mL of distilled water at 30 °C. After the samples were centrifuged (Biofuge 28 RS, Heraeus, Sepotech) for 15 min at 3000 rpm, the supernatant was removed and mixed with 10 mL of sulphuric acid (Merck, Darmstadt, Germany) and acetic anhydride (Sigma–Aldrich, Missouri, USA) (1:9) and placed in a thermostatic bath at 100 °C for 10 min (acetolysis). Then, the samples were centrifuged at 3000 rpm, the supernatant was removed and liquefied Kaiser's glycerol gelatine (Merck, Darmstadt, Germany) was added to the pollen grains sediment. Finally, the entire mixture was spread on a glass slide. Each analysis was carried out in triplicate.

At least, 500 pollen grains were counted per sample in order to confirm the botanical origin indicated on the label of the packaging. Pollen identification was based on the reference collection from the Laboratory of the School of Agriculture in Castelo Branco and some pollen atlases.

2.2. Physical–chemical analysis

A total of 10 parameters will be described: free acidity and pH, electrical conductivity, moisture content, water activity, ash content, total phenol content and colorimetric coordinates in CIELab space. Three measurements were done for each honey sample and parameter.

2.2.1. Free acidity and pH

Free acidity was determined by potentiometric titration and pH was estimated using a potentiometer (Radiometer PHM61, Hach Lange, USA), according to the harmonized methods of the European Honey Commission (Bogdanov, Martin, & Lullmann, 1997). Honey samples were homogenized in a water bath before analysis. 10 g of honey were dissolved in 75 mL of distilled water. The solution was titrated with 0.1 M NaOH (Sigma–Aldrich, Missouri, USA). Free acidity was expressed as milliequivalent of acid per kg of honey.

2.2.2. Electrical conductivity

The electrical conductivity was measured at 20 °C using a portable Conductivity meter Model WTW 315i (Weilheim, Germany). The instrument was calibrated using 0.01 M KCl (Merck, Darmstadt, Germany).

2.2.3. Moisture and water activity

The moisture content of the honey samples was determined in refractometer ATAGO-1T (Tokyo, Japan).

Water activity (aw) was evaluated in Rotronic Hygroskop DT equipment with a thermostatic WA-14TH probe (Zurich, Switzerland), where the sample is sealed in the cell apparatus, allowing the relative moisture in the honey sample and within the cell to remain in equilibrium. The sensor located at the top of the cell is constantly reading relative moisture, stabilizing when the variation is less than 0.02% HR/min to 0.02 °C/min.

2.2.4. Ash content

Total ash content was done according to the standard NP-1308 (1976), by incinerating honey samples in a muffle furnace (Nabertherm, Germany) at a temperature of 550 °C.

2.2.5. Phenol content

Total phenols were determined according to a modified Folin–Ciocalteu method (Meda, Lamien, Romito, Millogo, & Nacoulma, 2005; Waterhouse, 2003). Each honey sample (5 g) was diluted to 50 mL with distilled water and filtered through Whatman No. 1 paper. This solution (0.1 mL) was then mixed with 1.5 mL of water, 0.1 mL of Folin–Ciocalteu reagent (Panreac Quimica, SA, Barcelona, Spain) for 5 min and 0.3 mL of 20% (w/v) sodium carbonate (Na₂CO₃) (Panreac Quimica, SA, Barcelona, Spain) was then added to a total volume of 2 mL. After incubation at room temperature for 1 h, the absorbance of the reaction mixture was measured at 765 nm (JASCO 7800 UV/VIS spectrophotometer, Elnor, Germany). Gallic acid (Panreac Quimica, SA, Barcelona, Spain) (0–25 mg/L) was used as standard to produce the calibration curve.

The mean of three readings was used and the total phenol content was expressed in mg of gallic acid equivalents (GAE)/100 g of honey.

2.2.6. Color

The Color parameters (L^* , a^* , b^*) were measured in the CIELab system using a Minolta CR300 Chroma-meter with illuminant D65 (Osaka, Japan) by measuring lightness (L^*), redness (a^*), and yellowness intensities (b^*). Specimens were illuminated at 45°.

2.3. Artificial neural networks and data analysis

Artificial neural networks are machine learning techniques, where the aim is to program systems in such a way that they are able to automatically learn and improve with experience. Algorithms that discover knowledge from specific data and experience are needed, allowing computers to learn without being explicitly programmed.

Artificial neural networks try to copy the architecture of human brain and the connections of our biological neurons (Fausset, 1993; McCulloch & Pitts, 1943). It is composed by an interconnected group of artificial neurons that processes information with a connection approach to computation.

In an artificial neural network, an initial input layer (with the input data) is connected to one or more than one hidden layer, which is also connected to the output layer (the result needed). Each layer contains a certain number of nodes, and they are connected taking into account the weights that indicate the importance of each node.

Multilayer perceptron (MLP) is a particular case where each node is a perceptron, and is characterized by a back-propagation process (Bishop, 1995). The architecture is based on an acyclic diagram where each node is not backpropagated, having different weights and specific activation functions that define each particular neural network. The function $f: \mathbf{X} \subset \mathbb{R}^d \rightarrow \mathbf{Y} \subset \mathbb{R}^c$ is implemented (Heaton, 2012; Nagesh Kumar, Srinivasa Raju, & Sathist, 2004):

$$f(\mathbf{x}) = \phi(\psi(\mathbf{x})) \quad (1)$$

$$\phi: \mathbf{X} \subset \mathbb{R}^d \rightarrow \mathbf{T} \subset \mathbb{R}^p$$

$$\psi: \mathbf{T} \subset \mathbb{R}^p \rightarrow \mathbf{Y} \subset \mathbb{R}^c$$

where $\mathbf{X}$ is the input space with an activation function ψ , $\mathbf{T}$ represents the hidden space, ϕ is its activation function, and $\mathbf{Y}$ is the output space. The function implemented by the MLP can be expressed as (Bishop, 1995; Heaton, 2012; Mustafa, Rezaur, Saiedi, & Isa, 2012):

$$f(\mathbf{x}) = \sum_{j=1}^p \phi_j(c_j \psi(\mathbf{w}_j^T \mathbf{x} + \mathbf{w}_0) + \mathbf{c}_0) \quad (2)$$

where p represents the number of neurons contained in the hidden layer, c_j and c_0 are the weights of the hidden layer and $\mathbf{w}_j$ and $\mathbf{w}_0$ are the weights of the input layer. These weights are adjusted during the training process of the neural network using the appropriate algorithms (Bishop, 1995).

The operation of MLP can be summarised as follows: the artificial neurons, organized in layers (input, hidden and output), send the signals forward while the errors propagate backwards. This process propagates backwards the errors obtained during the training of the neural network, allowing the network to learn the analysed data, thus improving its performance.

In this research work, neural networks were used to predict the botanical origin from a number of physical and chemical parameters. Several tests were done varying the configuration of the input layer in order to optimize the neural network. A total of 14 different configurations were considered, with the following parameters in the input layer:

- Conf. 1: all the available variables are included as input data;
- Conf. 2: only the color parameters are considered (L^* , a^* , b^*);
- Conf. 3: the same as configuration 2, including ash content;
- Conf. 4: the same as configuration 2, including electrical conductivity;
- Conf. 5: the same as configuration 2, including phenol content;
- Conf. 6: the same as configuration 2, including pH;
- Conf. 7: color parameters, electrical conductivity, ash content and phenol content;
- Conf. 8: color parameters, electrical conductivity and phenol content;
- Conf. 9: color parameters a^* and b^* and electrical conductivity;
- Conf. 10: color parameters a^* and b^* , electrical conductivity and phenol content;
- Conf. 11: color parameters a^* and b^* , electrical conductivity and ash content;
- Conf. 12: color parameters a^* and b^* and ash content;
- Conf. 13: color parameters L^* and b^* , electrical conductivity, free acidity and phenol content;
- Conf. 14: color parameters a^* and b^* , electrical conductivity, ash content and phenol content.

The optimization of the neural network was done through three concepts: firstly, the number of neurons in the hidden layer was incremented to minimize the error of the prediction; the input variables of the hidden layer were also varied; and finally, the number of subsets in which the database was divided in the training stage of the neural network (k -fold) was considered due to its influence in the learning rate of the network. These concepts are explained in the next paragraphs.

Table 1
Dataset including all the samples. Three measurements were taken of each physical–chemical parameter.

Id	Botanical origin		No. samples	No. measurements	% Of specific pollen
A	Carob	<i>Ceratonia siliqua</i> L.	4	12	37
Am	Almond	<i>Prunus dulcis</i> (Miller) DA Webb	1	3	33
Az	Holm oak	<i>Quercus rotundifolia</i> Lam.	1	3	
C	Sweet chestnut	<i>Castanea sativa</i> Mill	1	3	60–65
E	Eucalyptus	<i>Eucalyptus globulus</i> Labill.	5	15	66–70
L	Orange	<i>Citrus</i> spp.	7	21	16–21
LV	Rosemary	<i>Rosmarinus officinalis</i> L.	1	3	20
R	Lavender	<i>Lavandula</i> spp.	12	36	16–18
Md	Strawberry trees	<i>Arbutus unedo</i> L.	1	3	9
T	Thyme	<i>Thymus vulgaris</i> L.	1	3	28
TR	Thyme and lavender	<i>Thymus vulgaris</i> L. y <i>Lavandula</i> spp.	1	3	–
U	Heather	<i>Erica</i> spp.	7	21	36–42
G	Sunflower	<i>Helianthus annuus</i> L.	1	3	45
N	Multifloral	–	6	18	–
		Total	49	147	

Table 2
Physical–chemical variation parameters for the different honey types.

Honey samples	Moisture (%)	pH	Color			Electrical conductivity (S/cm)	Ash (%)	Water activity	Free acidity (meq/kg)	Total phenol content (mg GAE/100 g)
			L*	a*	b*					
Carob	18.58 ± 1.17	3.66 ± 0.52	34.51 ± 6.76	3.35 ± 1.21	1.73 ± 1.49	644.89 ± 465.75	0.26 ± 0.21	0.59 ± 0.03	27.35 ± 5.44	42.55 ± 16.48
Almond	18.93 ± 0.12	3.40 ± 0.00	39.79 ± 0.01	2.50 ± 0.02	0.46 ± 0.00	273.67 ± 5.03	0.11 ± 0.01	0.59 ± 0.01	22.72 ± 0.11	22.32 ± 0.68
Holm oak	17.87 ± 0.12	4.40 ± 0.00	34.31 ± 0.01	4.38 ± 0.02	−3.40 ± 0.01	1113.67 ± 1.15	0.62 ± 0.01	0.59 ± 0.03	45.37 ± 0.40	81.03 ± 0.48
Sweet chestnut	17.20 ± 0.00	3.90 ± 0.00	36.83 ± 0.01	5.17 ± 0.01	−0.23 ± 0.01	640.33 ± 11.02	0.30 ± 0.01	0.56 ± 0.01	33.58 ± 0.43	67.56 ± 2.10
Eucalyptus	17.55 ± 0.61	3.40 ± 0.22	39.05 ± 0.53	2.95 ± 0.57	0.79 ± 0.42	430.00 ± 89.24	0.15 ± 0.05	0.57 ± 0.02	31.28 ± 8.53	26.32 ± 3.21
Orange	18.82 ± 0.39	3.50 ± 0.31	33.65 ± 6.70	1.95 ± 1.45	1.77 ± 2.43	322.17 ± 162.44	0.12 ± 0.11	0.60 ± 0.02	29.20 ± 7.47	37.15 ± 29.33
Rosemary	17.23 ± 0.12	3.23 ± 0.06	36.90 ± 0.02	3.18 ± 0.01	−1.01 ± 0.01	237.00 ± 1.73	0.07 ± 0.01	0.56 ± 0.02	25.72 ± 0.34	15.35 ± 0.12
Lavender	18.74 ± 1.48	3.81 ± 0.33	30.28 ± 4.95	0.74 ± 1.08	8.61 ± 4.89	325.68 ± 120.57	0.11 ± 0.06	0.61 ± 0.02	23.65 ± 7.69	23.24 ± 8.38
Strawberry trees	19.50 ± 0.50	4.20 ± 0.00	24.94 ± 0.00	2.23 ± 0.07	5.57 ± 0.02	863.33 ± 8.62	0.49 ± 0.01	0.66 ± 0.01	33.13 ± 0.72	44.48 ± 2.72
Thyme	18.23 ± 0.12	4.10 ± 0.00	38.64 ± 0.01	3.31 ± 0.02	1.63 ± 0.01	407.67 ± 1.53	0.17 ± 0.01	0.59 ± 0.01	23.77 ± 0.29	84.31 ± 1.22
Thyme and lavender	17.97 ± 0.06	3.00 ± 0.00	40.90 ± 0.01	2.03 ± 0.02	0.01 ± 0.01	156.83 ± 1.56	0.03 ± 0.02	0.58 ± 0.02	21.67 ± 0.10	11.79 ± 0.80
Heath	18.91 ± 0.59	3.91 ± 0.22	28.40 ± 6.68	3.73 ± 1.98	2.10 ± 3.05	640.52 ± 185.72	0.31 ± 0.10	0.60 ± 0.02	30.30 ± 6.13	60.64 ± 23.82
Sunflower	18.50 ± 0.00	3.90 ± 0.00	29.04 ± 0.44	0.73 ± 0.08	11.25 ± 0.04	484.33 ± 2.31	0.18 ± 0.00	0.62 ± 0.00	25.54 ± 0.60	17.27 ± 0.15
Multifloral	18.07 ± 0.94	3.54 ± 0.39	37.03 ± 2.61	4.30 ± 1.60	−1.52 ± 1.06	656.28 ± 474.32	0.34 ± 0.29	0.59 ± 0.03	37.75 ± 14.54	53.19 ± 28.62
Std of the mean values	0.68	0.39	4.81	1.30	3.98	263.78	0.17	0.03	6.58	24.32
Global mean value	18.47	3.69	33.36	2.57	2.99	479.41	0.21	0.59	29.06	39.36
Variation (%)	4	11	14	51	133	55	81	4	23	62

The errors in training and test stages were calculated, considering them as the main criterion in the optimization of the network. The number of neurons in the hidden layer was incremented until the overfitting was observed (when higher number of neurons did not imply better performance of the neural network).

Regarding the input parameters, their values were studied in order to determine the most influential ones, which were expected to be those whose values showed considerable differences for each botanical origin.

The third concept is directly related to the learning process of the neural network. In this study, this process was divided into two stages, training and testing, and a cross validation process was employed: the dataset was divided into k subsets, using $(k - 1)$ subsets for training and the remaining other for testing the performance of the network; the process was repeated k times so each time a different subset was used for testing. In a classification problem, the number of subsets (k -fold) is conditioned by the number of data of each class in the dataset, since the subsets must contain data from all the classes.

Bearing that in mind, the optimization of the neural network was performed in two stages:

1. Using the whole dataset. The fact that only one sample (3 measurements) was available from half of the honey types restricts the maximum k -fold for training the network (k -fold = 3);
2. Using data that included only the honey types with at least 2 samples (6 measurements). Then, the k -fold could be incremented.

The fact of having only 3 measurements from several botanical origins limits the process of training of the neural network, since the maximum number of subsets that can be made is 3. This means that 66.6% of the data is employed in the training stage (2 subsets from a total of 3), which is a low percentage compared to the usual 80–90%. In view of this limitation, a second stage of neural networks training was performed with the data of the honey types with more than 3 measurements: *Ceratonia siliqua* L. (4 samples, 12 meas.), *Eucalyptus globulus* Labill. (5 samples, 15 meas.), *Citrus* spp. (7 samples, 21 meas.), *Lavandula* spp. (12 samples, 36 meas.), *Erica* spp. (7 samples, 21 meas.) and multifloral (6 samples, 18 meas.).

Taking into account the minimum number of measurements taken from each honey type, the maximum k -fold that could be set was 12. Thus, the following k -folds were established: 6, 9 and 12.

Hence, in this second stage of the optimization of the neural network the procedure was the following:

1. A configuration of the input variables was set;
2. A k -fold was established: 6, 9 and 12;
3. The neural network was trained with an increasing number of neurons in the hidden layer until the overfitting was observed.

3. Results and discussion

3.1. Laboratory samples characterization

The results of the honey pollen percentage of specific pollen for each sample analysis are represented in Table 1. All honey samples considered for analyses and labelled as monofloral are in accordance with their characteristics (Table 1).

Table 2 includes the mean and standard deviation of the different parameters evaluated in the honey samples.

All measured parameters are in accordance with the Portuguese standard for honey (Decreto-Lei 214/2003, n.d.).

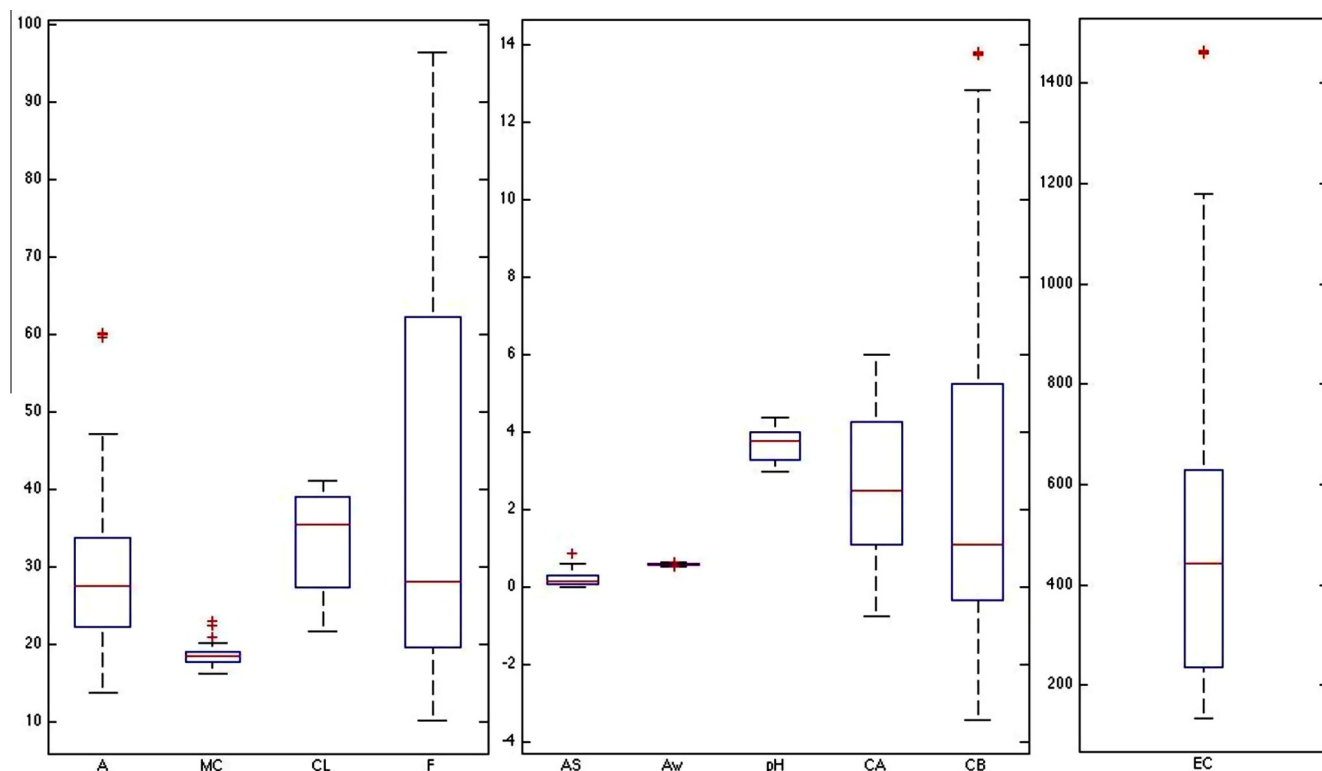


Fig. 1. Box-plot figures of the measured physical–chemical parameters. The central line indicates the median value, while the edges of the box represent the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted as individual crosses.

Table 3

Best results for each configuration of the input variables using the whole database (k -fold = 3) and the reduced database (k -fold = 6, 9 and 12). The error rates shown are calculated as the mean of the error rates for each individual honey type and for all the subsets. The number of neurons in the hidden layer is indicated in brackets and the best results are in bold.

Configuration	Whole database	Reduced database			
	k -Fold = 3	k -Fold = 6	k -Fold = 9	k -Fold = 12	
Conf. 1	47.62 (15)	41.32 (20)	7.51 (40)	15.42 (40)	
Conf. 2	38.10 (20)	38.72 (30)	6.89 (30)	14.05 (30)	
Conf. 3	48.98 (35)	34.34 (10)	10.80 (35)	15.80 (20)	
Conf. 4	51.02 (25)	44.83 (20)	6.48 (25)	15.09 (25)	
Conf. 5	50.34 (20)	49.56 (25)	8.13 (20)	15.65 (25)	
Conf. 6	51.02 (40)	45.74 (10)	7.51 (25)	18.93 (15)	
Conf. 7	48.30 (20)	45.21 (20)	4.22 (20)	13.33 (25)	
Conf. 8	45.58 (25)	41.38 (20)	7.00 (20)	14.38 (20)	
Conf. 9	40.14 (40)	34.05 (30)	5.25 (40)	12.68 (25)	
Conf. 10	44.90 (25)	44.02 (10)	4.42 (20)	12.29 (10)	
Conf. 11	47.62 (25)	37.81 (30)	4.42 (20)	11.96 (20)	
Conf. 12	48.98 (25)	38.72 (15)	8.95 (40)	16.46 (20)	
Conf. 13	48.98 (25)	38.44 (15)	7.82 (35)	13.72 (15)	
Conf. 14	45.58 (25)	42.23 (20)	8.33 (15)	12.95 (10)	

The average moisture content ranged from 17.2% to 19.5%, lower than the maximum moisture content value of pure floral honey, 23% for heather (*Calluna*) honeys and 20% in general. These results and the values observed for water activity suggests good honey conservation and storage. The water content of honey depends on the harvesting season, the degree of maturity reached in the hive, as well as from environmental factors (Acquarone, Buera, & Elizalde, 2007). Free acidity was lower than 50 meq/kg and may be explained by taking into account the presence of organic acids in equilibrium with their corresponding lactones, or internal esters, and some inorganic ions, such as phosphate (Finola, Lasagno, & Marioli, 2007; Ojeda de Rodríguez, Sulbarán

de Ferrer, Ferrer, & Rodríguez, 2004). According to Acquarone et al. (2007) this parameter can also be related with the botanical origin of honey.

The electrical conductivity of honey is directly related to the concentration of mineral, organic acids and proteins and must be lower than 800 $\mu\text{S}/\text{cm}$ (Acquarone et al., 2007). All the observed results are in accordance with the standards except for the holm oak honeys that showed higher electrical conductivity. However, in some studies for some samples of darker honey a higher electrical conductivity and ash content was observed, explained by the complexity of floral and geographic origin (Acquarone et al., 2007; Finola et al., 2007).

The range of values for ash content (0.1–0.6%) was within the limits proposed by the standards.

The values of pH and aw found guarantee the food safety of the honey samples.

Some authors reported that the phenol content could be characteristics for some monofloral honey namely heath, sunflower or orange honeys (Bertoncelj, Doberšek, Jamnik, & Golob, 2007). In this study the variation of phenol content is higher but the variability within each class does not suggest a possible correlation between them and the botanical origin.

The color of honey, in addition to the flavor and aroma, is one of the features that could allow a first classification of the honey floral origin, and can range from yellow to amber, red and dark amber to nearly black (Bertoncelj et al., 2007). The analysed honey types show a high variability of their color.

3.2. Neural networks optimization

In order to determine the most influential input parameters, we studied their values. Fig. 1 shows the box-plot of each variable. They are shown in three graphs because of their different range: the first graph ranges from 10 to 100 and includes free acidity

Table 4
Summary of results from the confusion matrixes. Percentage of correctly classified data.

		Botanical origin #class													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
Conf. 1	Kfold#1	25%	✓	✓	✓	40%	14%	✓	25%	0%	✓	✓	14%	✓	0%
	Kfold#2	✓	✓	✓	✓	✓	57%	✓	75%	✓	✓	✓	57%	✓	50%
	Kfold#3	25%	0%	✓	✓	40%	14%	✓	75%	✓	✓	✓	57%	✓	0%
Conf. 2	Kfold#1	50%	✓	✓	✓	40%	14%	✓	50%	✓	✓	✓	57%	✓	50%
	Kfold#2	✓	0%	✓	✓	40%	57%	✓	25%	✓	✓	✓	57%	✓	0%
	Kfold#3	✓	0%	0%	0%	0%	14%	0%	✓	0%	0%	0%	57%	0%	0%
Conf. 3	Kfold#1	50%	✓	✓	✓	40%	14%	✓	50%	✓	✓	✓	57%	0%	0%
	Kfold#2	✓	✓	✓	✓	40%	29%	0%	33%	✓	✓	0%	57%	0%	0%
	Kfold#3	25%	✓	✓	✓	40%	57%	✓	75%	✓	✓	✓	43%	✓	0%
Conf. 4	Kfold#1	25%	0%	0%	✓	60%	57%	✓	42%	✓	✓	✓	57%	✓	0%
	Kfold#2	✓	✓	✓	✓	40%	57%	✓	50%	✓	✓	✓	57%	0%	0%
	Kfold#3	25%	✓	✓	✓	40%	14%	✓	75%	✓	✓	✓	57%	✓	0%
Conf. 5	Kfold#1	50%	0%	✓	✓	40%	14%	✓	0%	✓	✓	✓	43%	✓	0%
	Kfold#2	✓	0%	✓	✓	40%	57%	✓	25%	✓	✓	✓	57%	✓	50%
	Kfold#3	✓	0%	✓	✓	✓	14%	✓	33%	✓	✓	✓	57%	✓	0%
Conf. 6	Kfold#1	0%	✓	✓	0%	40%	14%	✓	67%	✓	✓	✓	0%	✓	0%
	Kfold#2	✓	✓	✓	✓	40%	71%	✓	75%	✓	✓	✓	29%	✓	0%
	Kfold#3	25%	✓	✓	✓	40%	14%	✓	75%	✓	✓	✓	0%	✓	0%
Conf. 7	Kfold#1	25%	0%	✓	✓	✓	0%	0%	42%	0%	✓	✓	14%	✓	0%
	Kfold#2	✓	0%	✓	✓	40%	57%	✓	67%	✓	✓	✓	57%	✓	0%
	Kfold#3	0%	✓	✓	0%	40%	14%	✓	67%	✓	✓	✓	0%	✓	0%
Conf. 8	Kfold#1	75%	✓	✓	✓	✓	14%	✓	0%	✓	✓	✓	14%	✓	50%
	Kfold#2	✓	✓	✓	✓	40%	57%	✓	75%	✓	✓	✓	57%	✓	0%
	Kfold#3	25%	✓	✓	✓	40%	14%	✓	75%	✓	✓	✓	57%	✓	0%
Conf. 9	Kfold#1	25%	✓	✓	0%	80%	14%	✓	50%	✓	0%	✓	14%	✓	0%
	Kfold#2	✓	✓	✓	✓	40%	✓	✓	0%	✓	✓	✓	57%	✓	0%
	Kfold#3	75%	✓	✓	0%	40%	57%	✓	75%	✓	✓	✓	14%	✓	0%
Conf. 10	Kfold#1	25%	✓	✓	✓	✓	14%	✓	50%	✓	✓	✓	14%	✓	50%
	Kfold#2	✓	✓	✓	✓	40%	✓	✓	75%	✓	✓	✓	57%	✓	0%
	Kfold#3	50%	✓	✓	✓	40%	14%	✓	58%	✓	✓	✓	57%	✓	33%
Conf. 11	Kfold#1	25%	✓	✓	0%	40%	14%	✓	50%	✓	✓	0%	14%	✓	0%
	Kfold#2	✓	✓	✓	✓	40%	86%	✓	75%	✓	✓	✓	57%	✓	50%
	Kfold#3	25%	✓	✓	✓	40%	57%	✓	58%	0%	✓	✓	14%	✓	0%
Conf. 12	Kfold#1	25%	✓	✓	✓	80%	14%	✓	50%	✓	✓	✓	57%	✓	0%
	Kfold#2	✓	0%	✓	✓	40%	29%	✓	0%	✓	✓	✓	43%	✓	0%
	Kfold#3	25%	✓	✓	✓	40%	✓	✓	75%	0%	✓	✓	43%	✓	0%
Conf. 13	Kfold#1	25%	✓	✓	✓	40%	14%	✓	0%	✓	✓	✓	57%	✓	0%
	Kfold#2	✓	✓	✓	✓	✓	✓	✓	50%	✓	✓	✓	57%	✓	0%
	Kfold#3	25%	✓	✓	✓	40%	14%	✓	50%	✓	✓	✓	57%	✓	0%
Conf. 14	Kfold#1	25%	✓	✓	✓	✓	14%	✓	50%	✓	✓	✓	14%	✓	0%
	Kfold#2	✓	✓	✓	✓	✓	✓	✓	42%	✓	✓	✓	57%	✓	0%
	Kfold#3	25%	✓	✓	✓	40%	57%	✓	75%	✓	✓	✓	57%	✓	0%
Times all correctly classified		16	32	40	36	8	5	39	1	37	40	39	0	38	0
		38%	76%	95%	86%	19%	12%	93%	2%	88%	95%	93%	0%	90%	0%

✓ = 100% correctly classified.

(A), moisture content (MC), lightness of CIELab color coordinates (CL) and total phenol content (F); the second graph ranges from -4 to 14 and includes the ash content (AS), water activity (A_w), pH value (pH), redness intensity and yellowness intensity of CIELab color coordinates (CA and CB, respectively); finally, the third graph ranges from 150 to 1500 and contains the electrical conductivity values (EC).

The variation of the data was also studied taking into account the mean values of the parameters measured for each type of honey (Table 2). Calculating the standard deviation of the means of those values, and considering a percentage of variation (standard deviation of the means divided by the mean of all the measures of a certain physical–chemical parameter), it is possible to deduce which variables differ the most when the botanical origin varies.

Both Fig. 1 and Table 2 show that the electrical conductivity (EC), ash (AS), total phenol content (F) and colorimetric coordinates a^* and b^* (CA, CB) are the parameters whose values vary the most, and thus are expected to be worth being input parameters in the neural network. According to Table 2, the highest variations are the following: a^* color coordinate (51% variation), b^* color coordinate (133% variation), electrical conductivity (55% variation), ash content (81% variation) and total phenol content (62% variation). That is to say, these five properties differ the most from one honey type to another, so they are expected to be useful for the neural network learning process.

The following subsections contain the results of the two stages of optimization of the neural networks and their discussion.

3.2.1. Neural networks trained with the whole dataset

As it has been explained, the first neural networks were trained using all the available data, whose distribution was shown in Table 1. The k -fold used for the optimization of the network was 3, and the procedure was as follows: the input variables were selected according to the different configurations explained in Section 2.3 and the number of neurons in the hidden layer was increased until the error rate in test was minimum. Table 3 contains the best results for the 14 configurations. The minimum error rates achieved with each configuration are quite similar, ranging from 38.10% (Conf. 2) to 51.02% (Conf. 4 and 6).

We also obtained the confusion matrixes of each of the three iterations in order to study the accuracy of the predictions. We observed the following for the 14 configurations (Table 4): those classes with only one sample (three measurements) were correctly predicted (with no mistake) in 76–95% of the cases; however, the classes with more than one sample were hardly ever classified without errors (class 1 is correctly identified 16 times [38%], class 5: 8 times [19%], class 6: 5 times [12%], class 8: once [2%] and class 12 and class 14 are never correctly identified [0%]).

In view of the results, it seems that the number of measurements is influencing the results of the classification. Bearing this in mind, we decided to separate the data from botanical origins 1, 5, 6, 8, 12 and 14 and to construct new neural networks.

3.2.2. Neural networks trained with part of the database

Since the results obtained previously were quite similar for all the configurations of the input variables, we tested all of them with

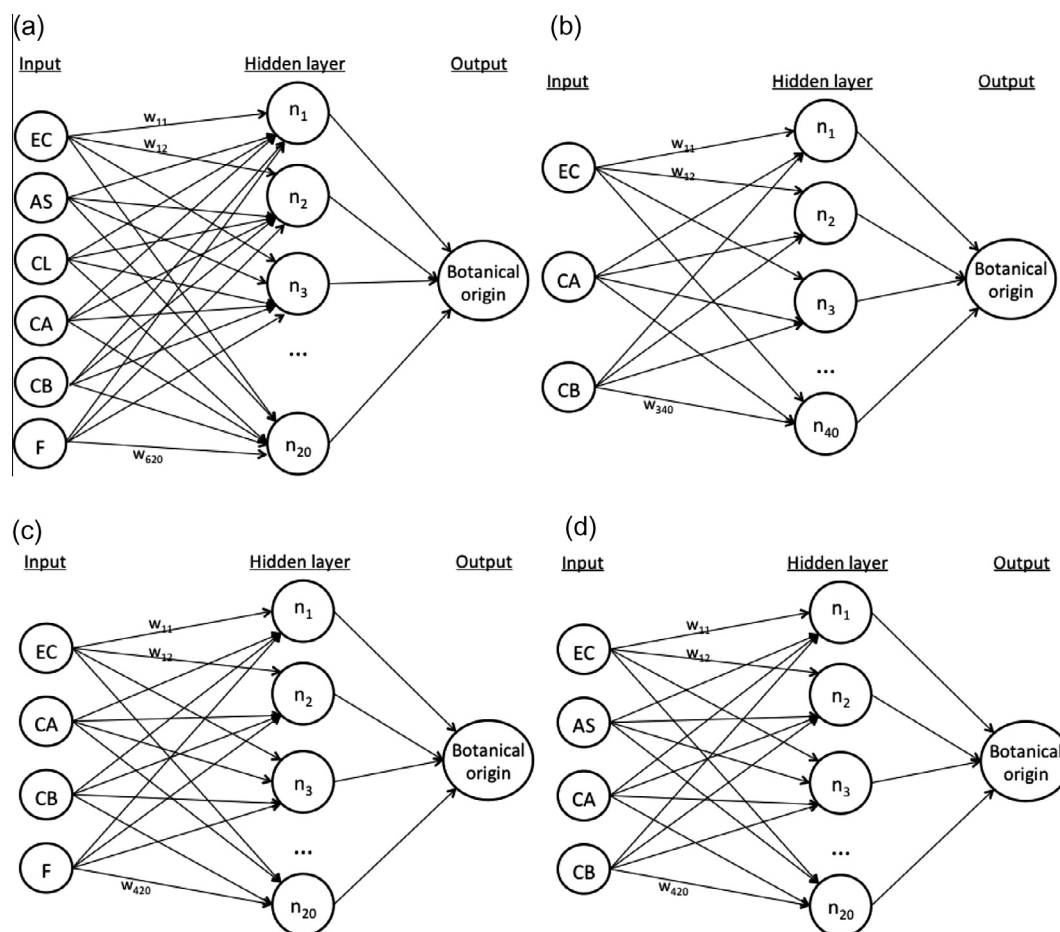


Fig. 2. Neural network of (a) configuration 7 with 20 neurons in the hidden layer, (b) configuration 9 with 40 neurons, (c) configuration 10 with 20 neurons and (d) configuration 11 with 20 neurons in the hidden layer.

the new reduced dataset (considering only Carob – class 1, Euca-lyptus – class 5, Orange – class 6, Lavender – class 8, Heath – class 12 and Multifloral – class 14 honeys). A minimum of 4 samples (12 measurements) were now available for each botanical class, so the k -fold was incremented in order to improve the behavior of the neural networks: 6, 9 and 12 subsets were used to train the networks. Table 3 contains the best results for each combination. It can be stated how the error rates are minimum when the number of subsets is 9 (8 subsets for training and 1 for testing).

We obtained and studied the confusion matrixes of the best configurations (conf. 7, conf. 9, conf. 10 and conf. 11, shown in Fig. 2) for the optimum case (k -fold = 9) and calculated the accuracy for each class and subset. We found clear similarities between the results of the four configurations:

- Class 14 is always correctly predicted, with 100% of the samples correctly identified regardless of the configuration and subset.
- Classes 6 and 12 show similar results: class 6 is well predicted in all cases except in one subset of configuration 10, where only 20% the samples are properly identified; regarding class 12, the classification is perfect except in one subset of configuration 9 that shows a 40% success. So these classes are always well predicted except in one of the subsets for one configuration.
- Finally, classes 1, 5 and 8 are misclassified in at least one of the subsets for every configuration. Class 1 is correctly classified in 8 out of 9 subsets for each configuration, and the same occurs with class 5. With respect to class 8, the results are slightly worse, since two subsets fail in the prediction with configuration 9, while the other configurations fail in one subset (as observed in the other cases).

So low error rates of about 5% are achieved with the configurations of the neural networks shown when the input data are divided into 9 subsets, as explained. These results suggest that neural networks are able to classify reasonably well the honey samples according to their botanical origin when the number of samples of each class allows a proper training process of the networks.

From the best input configurations, we can conclude that electrical conductivity and colorimetric coordinates a^* and b^* are essential to identify the class, since these three parameters are in the input layer of configurations 7, 9, 10 and 11. Hence, in order to establish a procedure for the systematic analysis at the laboratory, and to keep it as simple and quick as possible, electrical conductivity and colorimetric coordinates are the most convenient physical–chemical properties.

4. Conclusions

The results indicate that the botanical origin could be determined from the color and the electrical conductivity of the honey sample with low error rates (about 95% of the samples would be correctly classified). These results agree with previous works that pointed to the existing relationship between the color and the chemical composition of honey: Juszczak et al. (2009) and Acquarone et al. (2007), who also identified a correlation between electrical conductivity of the honey and botanical origin.

The performed tests suggest that the accuracy of the prediction of the botanical origin of these samples is improved when the database allows a more thorough training process of the neural networks. This fact makes us think that the error rates would be lowered with a more comprehensive database.

The proposed methodology suggest that it is possible to achieve a well-trained neural network able to classify different botanical origins of honey samples with electrical conductivity and the colorimetric coordinates a^* and b^* . Hence, they point the possibility

of innovation in the current melissopalynological procedures incorporating machine learning techniques to the analysis of the samples. Given the low error rates observed in this study, this methodology showed to be good and helpful to determine the honey botanical origin easily adapted to routine analysis of this product. However, further research work is needed in order to complete the knowledge of the neural network with more honey classes.

Acknowledgements

Centro de Estudos Florestais is a research unit funded by FCT (PEst-OE/AGR/UI0239-2014).

C. Iglesias is acknowledged to the Spanish Ministry of Education, Culture and Sports for FPU-12/02283 grant.

The authors would like to express their gratitude to Isabele Salvessa (IPCB Languages Centre) for the English revision and to Cecília Baptista and Conceição Vitorino for technical assistance in chemical determinations.

References

- Acquarone, C., Buera, P., & Elizalde, B. (2007). Pattern of pH and electrical conductivity upon honey dilution as a complementary tool for discriminating geographical origin of honeys. *Food Chemistry*, 101, 695–703. <http://dx.doi.org/10.1016/j.foodchem.2006.01.058>.
- Anjos, O., Campos, M. G., Ruiz, P. C., & Antunes, P. (2015). Application of FTIR-ATR spectroscopy to the quantification of sugar in honey. *Food Chemistry*, 169, 218–223. <http://dx.doi.org/10.1016/j.foodchem.2014.07.138>.
- Benedetti, S., Mannino, S., Sabatini, A. G., & Marazzan, G. L. (2004). Electronic nose and neural network use for the classification of honey. *Apidologie*, 35(4), 397–402. <http://dx.doi.org/10.1051/apido:2004025>.
- Bertoncelj, J., Doberšek, U., Jamnik, M., & Golob, T. (2007). Evaluation of the phenolic content, antioxidant activity and colour of Slovenian honey. *Food Chemistry*, 105, 822–828. <http://dx.doi.org/10.1016/j.foodchem.2007.01.060>.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.
- Bogdanov, S., Martin, P., & Lullmann, C. (1997). Harmonization methods of the European Honey Commission. *Apidologie* (Extra issue), 1–59.
- Caivano, J. L., & Buera, M. P. (2012). *Color in food: Technological and psychophysical aspects*. CRC Press Book.
- Cajka, T., Hajslova, J., Pudil, F., & Ridelova, K. (2009). Traceability of honey origin based on volatiles pattern processing by artificial neural networks. *Journal of Chromatography A*, 1216, 1458–1462. <http://dx.doi.org/10.1016/j.chroma.2008.12.066>.
- Cavaco, A. M., Miguel, G., Antunes, D., & Guerra, R. (2012). Determination of geographical and botanical origin of honey: From sensory evaluation to the state of the art of non-invasive technology. In G. Bondurand & H. Bosch (Eds.), *Honey: Production, consumption and health benefits* (pp. 1–40). New York (US): Nova Science Publishers.
- Cilla, M., Martínez, J., Peña, E., & Martínez, M. Á. (2012). Machine learning techniques as a helpful tool toward determination of plaque vulnerability. *IEEE Transactions on Bio-Medical Engineering*, 59(4), 1155–1161. <http://dx.doi.org/10.1109/TBME.2012.2185495>.
- Codex Alimentarius Commission FAO/OMS (2001). *Revised codex standard for honey, standards and standard methods*. Codex Alimentarius Commission FAO/OMS.
- Corvucci, F., Nobili, L., Melucci, D., & Grillenzoni, F.-V. (2015). The discrimination of honey origin using melissopalynology and Raman spectroscopy techniques coupled with multivariate analysis. *Food Chemistry*, 169, 297–304. <http://dx.doi.org/10.1016/j.foodchem.2014.07.122>.
- Decreto-Lei 214/2003, Pub. L. No. 216–18 de Setembro de 2003, 2003. Portugal: Diário da República.
- Dias, L. A., Peres, A. M., Vilas-Boas, M., Rocha, M. A., Estevinho, L., & Machado, A. A. S. C. (2008). An electronic tongue for honey classification. *Microchimica Acta*, 163(1–2), 97–102. <http://dx.doi.org/10.1007/s00604-007-0923-8>.
- Doner, L. W. (1977). The sugars of honey – A review. *Journal of the Science of Food and Agriculture*, 28(5), 443–456.
- Erdtman, G. (1960). The acetolysis method. A revised description. *Svensk Botanisk Tidskrift*, 39, 561–564.
- Escriche, I., Kadar, M., Domenech, E., & Gil-Sánchez, L. (2012). A potentiometric electronic tongue for the discrimination of honey according to the botanical origin. Comparison with traditional methodologies: Physicochemical parameters and volatile profile. *Journal of Food Engineering*, 109(3), 449–456. <http://dx.doi.org/10.1016/j.jfoodeng.2011.10.036>.
- European Union (2002). Council Directive 2001/110/EC of 20 December 2001 Relating to Honey. *Official Journal of the European Communities*.
- Fausset, L. V. (1993). *Fundamentals of neural networks: Architectures, algorithms and applications*. New York: Pearson.

- Finola, M. S., Lasagno, M. C., & Marioli, J. M. (2007). Microbiological and chemical characterization of honeys from central Argentina. *Food Chemistry*, 100(4), 1649–1653. <http://dx.doi.org/10.1016/j.foodchem.2005.12.046>.
- Gámbaro, A., Ares, G., Giménez, A., & Pahor, S. (2007). Preference mapping of color of Uruguayan honeys. *Journal of Sensory Studies*, 22(5), 507–519. <http://dx.doi.org/10.1111/j.1745-459X.2007.00125.x>.
- González-Miret, M. L., Terrab, A., Hernanz, D., Fernández-Recamales, M. A., & Heredia, F. J. (2005). Multivariate correlation between color and mineral composition of honeys and by their botanical origin. *Journal of Agricultural and Food Chemistry*, 53, 2574–2580. <http://dx.doi.org/10.1021/jf048207p>.
- González-Paramás, A. M., García-Villanova, R. J., Gómez Báñez, J. A., Sánchez Sánchez, J., & Ardanuy Albajar, R. (2007). Botanical origin of monovarietal dark honeys (from heather, holm oak, pyrenean oak and sweet chestnut) based on their chromatic characters and amino acid profiles. *European Food Research and Technology*, 226, 87–92. <http://dx.doi.org/10.1007/s00217-006-0512-9>.
- Heaton, J. (2012). *Introduction to the math of neural networks*. New York: Heaton Research.
- Iglesias, C., Martínez Torres, J., García Nieto, P. J., Alonso Fernández, J. R., Díaz Muñiz, C., Piñeiro, J. I., et al. (2014). Turbidity prediction in a River Basin by using artificial neural networks: A case study in Northern Spain. *Water Resources Management*, 28(2), 319–331. <http://dx.doi.org/10.1007/s11269-013-0487-9>.
- Juszczak, L., Socha, R., Rożnowski, J., Fortuna, T., & Nalepka, K. (2009). Physicochemical properties and quality parameters of herbhoney. *Food Chemistry*, 113(2), 538–542. <http://dx.doi.org/10.1016/j.foodchem.2008.07.098>.
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5, 115–133. <http://dx.doi.org/10.1007/BF02478259>.
- Meda, A., Lamien, C. E., Romito, M., Millogo, J., & Nacoulma, O. G. (2005). Determination of the total phenolic, flavonoid and proline contents in Burkina Faso honey, as well as their radical scavenging activity. *Food Chemistry*, 91(3), 571–577. <http://dx.doi.org/10.1016/j.foodchem.2004.10.006>.
- Mesallam, A. S., & El-Shaarawy, M. I. (1987). Quality attributes of honey in Saudi Arabia. *Food Chemistry*, 25(1), 1–11.
- Mustafa, M. R., Rezaaur, R. B., Saiedi, S., & Isa, M. H. (2012). River suspended sediment prediction using various multilayer perceptron neural network training algorithms—A case study in Malaysia. *Water Resources Management*, 26(7), 1879–1897. <http://dx.doi.org/10.1007/s11269-012-9992-5>.
- Nagesh Kumar, D., Srinivasa Raju, K., & Sathist, T. (2004). River flow forecasting using recurrent neural networks. *Water Resources Management*, 18(2), 143–161. <http://dx.doi.org/10.1023/B:WARM.0000024727.94701.12>.
- NP-1308:1976, Ed1, Mel. Determinação do teor de cinza.
- Ojeda de Rodríguez, G., Sulbarán de Ferrer, B., Ferrer, A., & Rodríguez, B. (2004). Characterization of honey produced in Venezuela. *Food Chemistry*, 84(4), 499–502. [http://dx.doi.org/10.1016/S0308-8146\(02\)00517-4](http://dx.doi.org/10.1016/S0308-8146(02)00517-4).
- Ouchemoukh, S., Schweitzer, P., Bachir Bey, M., Djoudad-Kadjji, H., & Louaileche, H. (2010). HPLC sugar profiles of Algerian honeys. *Food Chemistry*, 121, 561–568. <http://dx.doi.org/10.1016/j.foodchem.2009.12.047>.
- Shafiee, S., Minaei, S., Moghaddam-Charkari, N., & Barzegar, M. (2014). Honey characterization using computer vision system and artificial neural networks. *Food Chemistry*, 159, 143–150. <http://dx.doi.org/10.1016/j.foodchem.2014.02.136>.
- Terrab, A., Diez, M. J., & Heredia, F. J. (2002). Chromatic characterisation of Moroccan honeys by diffuse reflectance and tristimulus colorimetry – Non-uniform and uniform colour spaces. *Food Science and Technology International*, 8(4), 189–195. <http://dx.doi.org/10.1177/1082013202008004556>.
- Tuberoso, C. I. G., Jerkovic, I., Sarais, G., Congiu, F., Marijanovic, Z., & Kus, P. M. (2014). Color evaluation of seventeen European unifloral honey types by means of spectrophotometrically determined CIE L* – C*ab – h ab chromaticity coordinates. *Food Chemistry*, 145, 284–291. <http://dx.doi.org/10.1016/j.foodchem.2013.08.032>.
- Ulloa, P. A., Guerra, R., Cavaco, A. M., Rosa da Costa, A. M., Figueira, A. C., & Brigas, A. F. (2013). Determination of the botanical origin of honey by sensor fusion of impedance e-tongue and optical spectroscopy. *Computers and Electronics in Agriculture*, 94, 1–11. <http://dx.doi.org/10.1016/j.compag.2013.03.001>.
- Waterhouse, A. L. (2003). *Determination of total phenolics. Current protocols in food analytical chemistry*. John Wiley and Sons, Inc..
- Wilczyńska, A. (2014). Effect of filtration on colour, antioxidant activity and total phenolics of honey. *LWT – Food Science and Technology*, 57(2), 767–774. <http://dx.doi.org/10.1016/j.lwt.2014.01.034>.