Food Chemistry 267 (2018) 36–42

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Computational intelligence applied to discriminate bee pollen quality and botanical origin

Paulo J.S. Gonçalves [a,b], Letícia M. Estevinho [c,d], Ana Paula Pereira [c], João M.C. Sousa [b], Ofélia Anjos [a,e,*]

[a] Instituto Politécnico de Castelo Branco, 6000-084 Castelo Branco, Portugal
[b] IDMEC, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal
[c] Agricultural College of Bragança, Polytechnic Institute of Bragança, 5301-855 Bragança, Portugal
[d] Centre of Molecular and Environmental Biology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[e] Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade Lisboa, 1349-017 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 9 January 2017
Received in revised form 21 April 2017
Accepted 2 June 2017
Available online 3 June 2017

Keywords:
Bee pollen
Physical–chemical parameters
Botanical origin
Neural networks
Fuzzy modelling
Support vector machines

ABSTRACT

The aim of this work was to develop computational intelligence models based on neural networks (NN), fuzzy models (FM), and support vector machines (SVM) to predict physicochemical composition of bee pollen mixture given their botanical origin. To obtain the predominant plant genus of pollen (was the output variable), based on physicochemical composition (were the input variables of the predictive model), prediction models were learned from data. For the inverse case study, input/output variables were swapped.

The probabilistic NN prediction model obtained 98.4% of correct classification of the predominant plant genus of pollen. To obtain the secondary and tertiary plant genus of pollen, the results present a lower accuracy.

To predict the physicochemical characteristic of a mixture of bee pollen, given their botanical origin, fuzzy models proven the best results with small prediction errors, and variability lower than 10%.

1. Introduction

Bee pollen has a variable chemical composition, being the major compounds proteins, carbohydrates, sugar, lipids, fibres, vitamins and minerals (Arruda, Pereira, de Freitas, Barth, & de Almeida-Muradian, 2013; Campos et al., 2016; Estevinho, Rodrigues, Pereira, & Feás, 2012; Yang et al., 2013) and depends of the floral origin (Morais, Moreira, Feás, & Estevinho, 2011). According Campos, Olena, and Anjos (2016) bee pollen contain all the essential amino acids needed for the human organism.

Bee pollen is consumed throughout the world given its nutritional and therapeutic value (Amâcio, Serrano, Anjos, & Campos, 2014), even though the labeling of this product is particularly difficult due to the higher variability. Consumers acquire bee pollen in packaging containing different weight and the composition could be ranging between some values according the percentage of the different plant sources or the number and species present in a sample (Feás, Vázquez-Tato, Estevinho, Seijas, & Iglesias, 2012).

This natural product possesses promising pharmacological properties namely antioxidant, antimicrobial, antiviral, anti-inflammatory, antimutagenic, hepatoprotective and antiallergenic, that have been linked in great extent to the chemical composition, particularly to the content in phenolic compounds (Campos et al., 2008; Feas, Vazquez-Tato, Estevinho, Seijas, & Iglesias, 2012; Pascoal, Rodrigues, Teixeira, Feás, & Estevinho, 2014; Salles et al., 2014).

Computational intelligence models based on NN, FM and SVM focuses on the biological learning process and try to emulate it through algorithms able to learn from given data and provide new results (Anjos et al., 2015). These techniques have been claimed to be an effective tool for modelling and predicting different parameters among which the rheological behavior of honey (Ramzi, Kashaninejad, Salehi, Sadeghi Mahoonak, & Ali Razavi, 2015). Also, these tools were also applied in the identification of the relationship between honeys' chemical and electrical parameters (Pentoś, Łuczycka, & Wróbel, 2014).

Anjos et al. (2015) using neural network, conclude that the botanical origin of honey can be reliably and quickly known from the colorimetric information and the electrical conductivity of honey. Shafiee, Minaei, Moghaddam-Charkari, Ghasemi-Varnamkhasti, and Barzegar (2013) where estimated ash content,

* Corresponding author at: Instituto Politécnico de Castelo Branco, 6000-084 Castelo Branco, Portugal.
E-mail address: ofelia@ipcb.pt (O. Anjos).

http://dx.doi.org/10.1016/j.foodchem.2017.06.014
0308-8146/

antioxidant activity and total phenolic content from honey, using neural networks and color information. In this study, machine vision was used for the honey characterization and shows the importance of their applicability by the industry. According the same authors, there technique may be skillful by employing image analysis for honey color quantification.

At the best knowledge of the authors, no such prediction models exist in the literature. As such, the framework proposed in this paper is novel and not tackled before. Computational intelligence approaches were only applied to pollen in related fields. Some examples are for the identification of pollen texture (Li & Flenley, 1999), the classification of pollen grains (Dhawale, Tidke, & Dudul, 2013), to forecast airbone pollen concentration (Ranzi, Lauriola, Marletto, & Zinoni, 2003). The aim of this work is to develop a model based on neural networks (NN), fuzzy models (FM), and support vector machines (SVM) to predict physicochemical composition of a bee pollen mixture given their botanical origin and help in an easier labeling of this product.

2. Materials and methods

2.1. Sample characterization

The pollen samples (n = 210) were collected directly from beekeepers in the spring of 2015 at different regions in Portugal and stored in the dark at room temperature (±15 °C) until further analysis.

The percentage of pollen grains belonging to each botanical family was determined based on the observation of 300 to 400 pollen grains (mean value ± standard deviation; 335 ± 29%) in slides prepared by washing the pollen load in 50% ethanol and using glycerin and paraffin for permanent preparations. The observation of pollens was carried out with a Leitz Diaplan microscope (Leitz Messtechnik GmbH, Wetzlar, Germany) at ×400 and ×1000. The reference collection of the Agricultural College of Bragança (ESA-IPB) and different pollen morphology guides were used for the recognition of the pollen grains. The classification in frequency classes was performed as follows: predominant pollen (>45% of a specific pollen type), secondary pollen (15–45%), important minor pollen (3–15%) and minor pollen (<3%); according to the methodology proposed by Morais et al. (2011) and their distribution were plotted in Table 1. The samples were characterized by a higher amount of *Cistus* spp. as predominant genus in a higher number of samples. Others important genus that was representative as predominant pollen is *Eucalyptus* spp., *Echium* spp, and *Cytisus* spp. For the secondary pollen some genus are important in the samples namely *Castanea* spp., *Eucalyptus* spp., *Rubus* spp., *Cytisus* spp., *Erica* spp., *Genista* spp. and *Lavandula* spp..

2.2. Chemical analysis

Several analytical parameters were determined in the bee pollen samples, namely: moisture content, ph, water activity, ash, carbohydrates, reducing sugars, lipids, proteins, fiber content, total polyphenol content, flavonoids content and antioxidant capacity assessed by calculating the 2,2-diphenyl-1-picrylhydrazyl (DPPH - EC_{50}) free radical scavenging activity:

1) Moisture of the pollen samples was determined according to the AOAC procedures (AOAC, 1995);
2) pH was measured in the aqueous phase, obtained after mixing 10 g of pollen in 75 mL of distilled water, using a digital pH Meter (pH 526 Multical, WTW, Weilheim, Germany);
3) Ash content was determined by gravimetry after ignition at 600 ± 15 °C, as reported by Carpes, Begnini, de Alencar, and Masson (2007);
4) Reducing sugar were determined according the methodology proposed by (Estevinho et al., 2012). 60 mg of each pollen pellet was dissolved in a H_2SO_4 solution (10 mL, 1.5 M). The solutions were heated in a water bath (100 °C) for 20 min, neutralized with 12 mL NaOH (10%, w/v), filtered and the volume of the flask (60 mL) was completed with distilled water. The quantification of reducing sugars was performed spectrophotometrically at 540 nm using a spectrophotometer (UV–VIS spectrometry Unicam Hekios, UK);
5) For the determination of the total lipid content, two grams of pollen were macerated with anhydrous Na_2SO_4, and extracted with n-hexane for about 4 h in the Soxhlet apparatus (Bárbara et al., 2015);
6) Protein content was determined from the total nitrogen using the conversion factor 6.25 (N × 6.25), using the Kjeldahl method (230-Hjeltec Analyzer, Foss Tecator, Höganäs, Sweden);
7) The fiber percentage was determined by the official method recommended by (AOAC, 1995);
8) The total phenolic content (TPC) of the extracts was determined using the Folin–Ciocalteu method as described by Moreira, Dias, Pereira, and Estevinho (2008) and expressed as mg of Galic Acid equivalents per g of bee pollen (GAE/g pollen);
9) Flavonoids contents determination the aluminium chloride method was used. Total flavonoids content were expressed as mg of Quercetin equivalents per g of bee pollen (QE/g pollen);
10) The evaluation of the free radical blocking effect of DPPH (2,2-diphenyl-1-picrylhydrazyl) was performed according

Table 1
Frequency of each plant genus as predominant pollen, secondary pollen, important minor pollen and minor pollen in the 210 pollen samples mixture analysed.

Genus	Predominant pollen (>45%)	Secondary pollen (15–45%)	Important minor pollen (3–15%)	Minor pollen (<3%)
Eucalyptus spp.	13	24	21	33
Prunus spp.	8	4	7	14
Rubus spp.	11	11	11	8
Castanea spp.	21	34	23	15
Trifolium spp.	0	6	11	27
Cistus spp.	56	17	12	16
Echium spp.	24	20	14	16
Kiwi spp.	2	0	4	11
Leontondon spp.	3	2	20	23
Quercus spp.	0	6	14	21
Erica spp.	4	7	9	13
Thymus spp.	2	0	11	2
Genista spp.	0	6	6	6
Lavandula spp.	0	4	13	0
Cytisus spp.	14	23	11	12

to the methodology described by Moreira et al. (2008). The percentage of the blocking effect of DPPH was calculated by the following equation: Radical Scavenging Activity (%) = [(ADPPH − AS)/ADPPH] × 100, where ADPPH is the absorbance of DPPH solution and, AS is the absorbance of the solution when the sample extract has been added at a particular level. The concentration of the extract that induced a 50% inhibition (EC_{50}) was calculated from the graphic of the percentage effect of eliminating radicals as a function of the concentration of the sample extract solution.

The percentage of Carbohydrates or the Energy (need in the labeling) were not considered in this work because are dependent from the other ones, i.e., are evaluated by applying a equation and not directly measured. As such, these values are autocorrelated with the other input/output data, which gives no benefit to the learning methods. Will only increase the complexity and computational time required for the learned methods training and application.

2.3. Computational intelligence based models and data analysis

The proposed twofold methodology, to obtain estimation and classification models, of pollen sources and its physical-chemical properties, was based in three different types of classifiers/estimators: fuzzy models, neural networks and support vector machines. The workflow is depicted in Fig. 1, and have three main steps: data preparation, classification/estimation, and finally the results analysis.

In the general case, data is organized in inputs and outputs of the system to be modelled. The input data is defined by a vector, $\boldsymbol{x} = [x_1 x_2 \dots x_n]T$, where n, is the number of inputs. The output data is defined by a vector $\boldsymbol{y} = [y_1 y_2 \dots y_m]T$, where m, is the number of outputs. From the learning approaches depicted in the following sub-sections, several models, $\boldsymbol{y} = \boldsymbol{F}(\boldsymbol{x})$, will be obtained to solve the stated classification and regression problems. Depending on the learning approaches, Multiple Input Multiple Output (MIMO) models, or several Multiple Input Single Output (MISO) models, will be obtained for mapping the plant genus to the correspondent physicochemical properties, and vice versa, of bee pollen.

2.3.1. Data preparation

Data preparation encompasses normalization, and data separation for training and validation, to obtain a proper model. Usually, a classifier or regression model is generated using part of the available data. If we use the whole data set available, it is very easy to get exceptional results. However, these results would likely not generalize to new examples. Computational Intelligence researchers, e.g., (Gonçalves, 2013), have developed training techniques that reduce the likelihood of overfitting to the training data.

Cross validation (Geisser, 1993) is a model evaluation method that does not use the entire data set when training a learner. Some of the data is removed before training begins. Then when training is done, the data that was removed can be used to test the performance of the learned model on new data. This is the basic idea for a whole class of model evaluation methods called cross validation.

After preparing the data for classification, taking in mind the previous paragraphs, the classification algorithms can be applied.

2.3.2. Fuzzy models

From the modelling techniques based on soft computing, fuzzy modelling (Sousa & Kaymak, 2002) is one of the most appealing. If no a priori knowledge is available, the rules and membership functions can be directly extracted from process measurement. Fuzzy models can provide a transparent description of the system, which can reflect the nonlinearites of the system. This paper uses Takagi-Sugeno fuzzy models (Takagi & Sugeno, 1985) where the consequents are crisp functions of the antecedent variables.

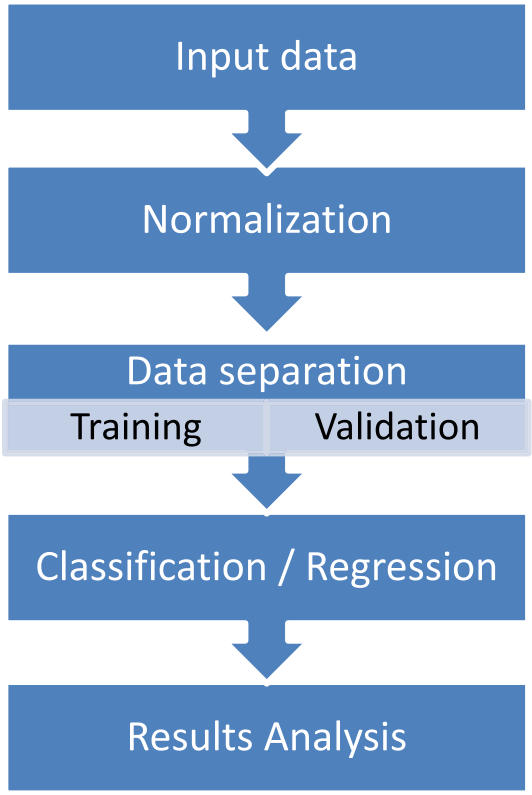


Fig. 1. General Workflow proposed in the paper, to obtain the learned models, from gathered data: plant genus and physicochemical parameters. Follows a normalization of the data, and after it is split for training the models and validate them. The classification and regression models are then obtained to be classify the main pollen and to estimate the physicochemical parameters.

Different classes of fuzzy clustering algorithms can be used to approximate a set of data by local models. From the available clustering algorithms, only Fuzzy C-Means (FCM) (Bezdek, 1981), and subtractive clustering (SC), (Chiu, 1994), were used in this study.

Fuzzy Inference systems are if-then rule based, and each rule, which is associated to the number of data clusters, K. Rules, R_i, have antecedents, associated to fuzzy sets, A_{i1}, and consequents, B_i, Eq. (1).

$$\begin{aligned} R_i: \quad & \textit{If } x_1 \textit{ is } A_{i1} \textit{ and} \dots \textit{ and } x_n \textit{ is } A_{in} \\ & \textit{then } y_i = B_i, \quad i = 1, 2, \dots, K. \end{aligned} \tag{1}$$

The number of rules, K, and the antecedent fuzzy sets A_{ij} are determined by means of fuzzy clustering in the product space of the inputs and the outputs (Castilho, Goncalves, Pinto, & Serafim, 2007). To obtain each estimated output, $\hat{y}$, Eqs. (2) and (3) were used to average the contribution of each rule, where β_i is the degree of activation of each rule, and $\mu_{A_{ij}}(x_j)$ is the membership function of each fuzzy set A_{ij}.

$$\hat{y} = \frac{\sum_{i=1}^{K} \beta_i y_i}{\sum_{i=1}^{K} w_i \beta_i}, \tag{2}$$

$$\beta_i = \Pi_{j=1}^{n} \mu_{A_{ij}}(x_j), \quad i = 1, 2, \dots, K, \tag{3}$$

2.3.3. Neural networks

Neural Networks (NN) (McCulloch & Pitts, 1943) are based on the interconnection between a set of simple processing units

(neurons). Each of these neurons contains a linear or nonlinear transformation function. The connections between the neurons have an associated weight that must be trained in order to adjust the performance of the network to the purpose of its use.

NN are flexible classification systems that are easily trained using backpropagation. By repeatedly showing a neural network inputs classified into groups, the network can be trained to discern the criteria used to classify, and it can do so in a generalized manner allowing successful classification of new inputs not used during training (Zhang, 2000). Each NN is characterized by the number of its layers, the units that compose each one of these layers, the interconnections between the units composing each layer and the ones that compose the following layer and all the associated weights.

The probabilistic neural network was first presented by Specht (1990) and is a type of feedforward neural network, obtained from the Bayes Net and the Kernel Fisher discriminant analysis, as presented in the seminal paper. This NN is specially designed for classification problems, and in this work was used to predict the plant genus from the physicochemical properties of the bee pollen. The network architecture have three layers: the input, the radial basis, and the competitive layers. The input layer calculates the distances, by a hidden neuron, from the input data to the training data set. These values are then sent to the radial basis functions, with a given spread, and a vector of probabilities is obtained. In the competitive layer, is chosen the large probability to define the main plant genus class. Fig. 2 depicts the PNN architecture, previously presented, and the main mathematical equations, showing the number of inputs, outputs and radial based functions obtained for the network trained with the data studied in this paper.

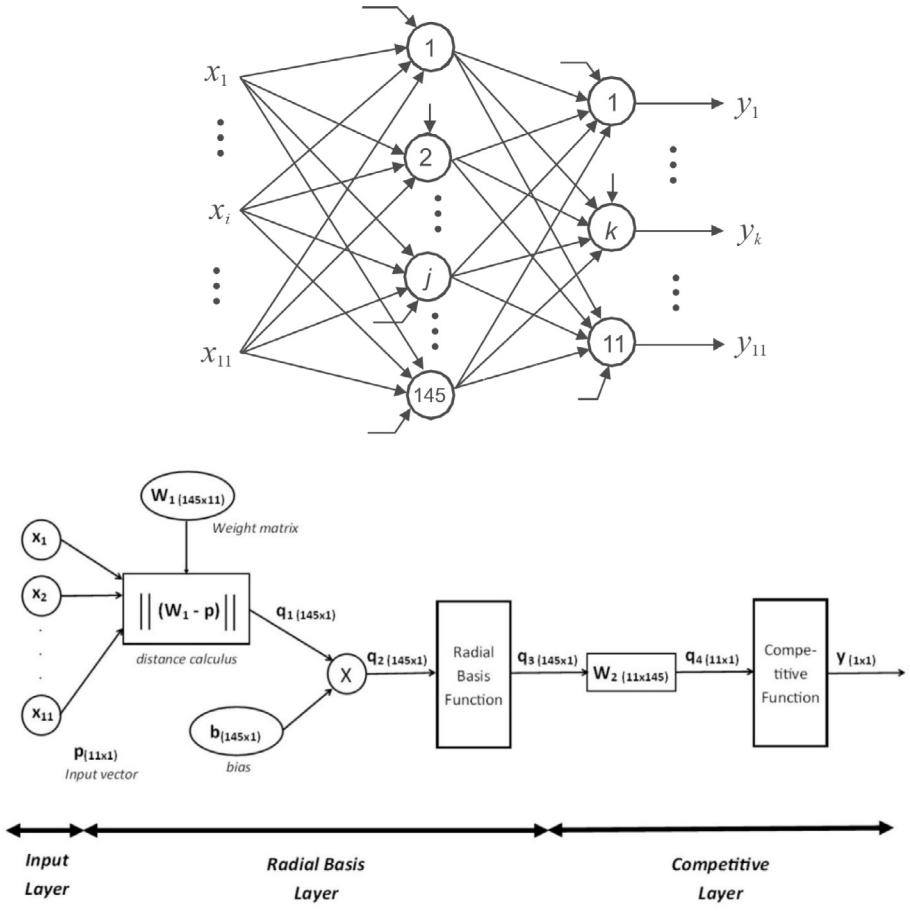
2.3.4. Support vector machines

The Support Vector Machine (SVM) (Vapnik, 1998), maps an input vector into a high-dimensional descriptor space through some nonlinear mapping, chosen a priori. In this space, an optimal separating hyperplane is constructed. SVM classification was performed using linear and quadratic penalization of misclassified examples, where the penalization coefficients can be different for each example, SVM classification with nearest point algorithm. SVM classification can use the strategies one against all, one against one and also large scale SVM. All theoretical candidates for the classification task at hand.

SVM regression was applied for the estimation case, and the linear epsilon loss insensitive optimization technique was used, where the Gaussian kernel's bandwidth have to be set, along with the Lagrangian multipliers (Vapnik, 1998). The purpose of the method is to obtain the linear function,

$$f(x) = x^T \cdot \beta + b \tag{4}$$

This function is obtained by formulating a convex optimization problem to find $f(x)$ with minimal value for $\beta^T \cdot \beta$, with the following stopping criteria for the residuals: $\forall_n : |y_n - (x^T \cdot \beta + b)| \leqslant \varepsilon$.



$$q_4(p) = W_2 \cdot RBF\left(\|W_1 - p\|.b\right) \qquad y(p) = argmax\{(q_4)_i(p)\},\ i = 1,2,..,11$$

Fig. 2. Probabilistic Neural Network architecture and implementation scheme, depicting 145 neurons in the radial basis layer, the weights W1 and W2, the bias b, for 11 inputs and the output class (with the maximum probability out of the 11 possible classifications).

Following (Vapnik, 1998) the goal is to minimize the following Lagrangian equation, where nonnegative multipliers, and α_k^*, were introduced for each data sample k, from the N available.

$$L(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i^T \cdot x_j + \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*)x_i^T \cdot x_j + \sum_{i=1}^{N} y_i(\alpha_i^* - \alpha_i) \quad (5)$$

Constrained with:

$$\sum_{i=1}^{N}(\alpha_n - \alpha_n^*) = 0$$

$$\forall_n : 0 \leqslant \alpha_k \leqslant C$$

$$\forall_n : 0 \leqslant \alpha_k^* \leqslant C$$

2.3.5. Performance indexes for results analysis

In this sub-section are presented two types of performance indexes, for the twofold methodology, i.e., related to the classification and regression problems. These indexes are needed to properly examine the data and the learned fuzzy models, neural networks or support vector machines, it is necessary to carefully evaluate the quality of the developed models.

The models were developed using Matlab, while using the CLAP platform (Gonçalves, 2013) and the toolboxes therein for classification and/or regression. The toolbox is freely available by request to the authors.

Accuracy index was used as the relevant index for the assessment of the learned models, for the main/principal, secondary and tertiary classes of pollens. The confusion matrix (Kohavi & John, 1997) was used to present the results for the assessment of the classification of the main/principal class of pollen.

In the regression case, and for the assessment of the results obtained, three indexes were used that are complementary, i.e., the R^2 index and the F-test, and also the Root Mean Square (RMSE). R^2 and F-test indexes are related to the variation of the variables, the dispersion test when equal to 1, rejects the null hypothesis at the default 5% significance level.

3. Results and discussion

The results in this section were obtained from estimated computational intelligence models, based in neural networks, fuzzy models and support vector machines. Using these three methodologies, several studies were carried out by the authors amongst possible methods within each class. In the next sub-sections, tables are presented for the best models obtained, although several tests were made with other methods, presented in CLAP, the Classification Platform (Gonçalves, 2013).

The database used included 210 samples of data, and each sample comprised information regarding the eleven sources of pollen and the eleven physicochemical parameters. In Table 1 is depicted the input/output data, i.e., the frequency of each plant genus as predominant pollen, secondary pollen, important minor pollen and minor pollen in the 210 pollen samples mixture analysed.

The values of physicochemical were similar to the reports available in the literature (Arruda et al., 2013; Campos et al., 2008, 2016; Estevinho et al., 2012).

3.1. Pollen classification

The pollen classification approach was designed in two steps. In the first step, with results presented in the first two rows of Table 2 and the confusion matrix of Fig. 3, the goal was to obtain a model that allowed to classify all the pollen sources by introducing as input the results of the physicochemical analysis. In other words, to estimate the principal components/sources of pollen given its physicochemical characteristics. The second step, with results presented in the last three rows of Table 2, aimed to obtain the principal component of the pollen source, in order to enhance the results obtained by applying the previous approach.

Table 2
Results from the classification of the plant genus given the bee pollen physicochemical parameters. First two rows the three predominant bee pollen plant genus. Last three results for classification of the predominant bee pollen plant genus.

Method	% Accuracy		
	Main pollen plant genus	Secondary pollen plant genus	Tertiary pollen plant genus
NN – FFwBRP	88.10	51.90	24.76
Fuzzy – Mamdani – FCM	92.06	73.02	46.03
Fuzzy – Sugeno – SC	90.32	–	–
NN – pnn	98.39	–	–
SVM – oneaone	90.32	–	–

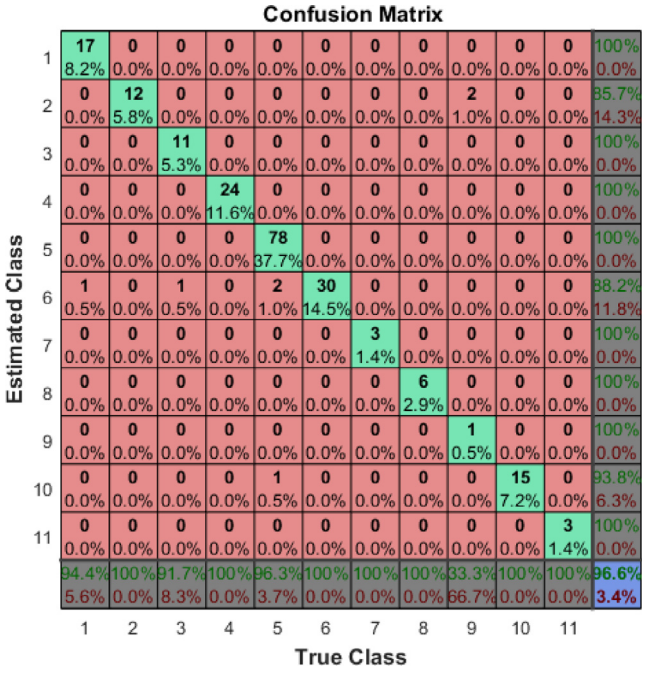


Fig. 3. Confusion matrix for a learned model based on a feedforward neural network with back-propagation training.

For the first step two approaches were investigated to obtain the classification model, based in neural networks and fuzzy models. The first used a feed-forward neural network, trained using back-propagation (BRP) and with 40 clusters. The second, i.e., fuzzy inference system, was obtained with FCM clustering and mamdani consequents (Fuzzy – Mamdani – FCM) to obtain the outputs. For the training and validation steps, the data was normalized and split with 70% for training and 30% for validation. A 5-fold cross validation approach was used, i.e., the data was randomly divided in training and validation, for five experiments, and the results are depicted in the first two rows of Table 2. It is observed that the fuzzy inference system performs better than NN, for all the three main components of pollen. The fuzzy estimated model achieves 92.06% accuracy in determining the principal component of pollen, based in its physicochemical characteristics. This result is observed for the estimation of the secondary and tertiary pollen components.

For the second step of the classification process, as depicted in rows three to six of Table 2, the effort to obtain better results for the estimation of the main pollen component was reward. Fig. 3 presents the results of the confusion matrix for a learned model based on a feedforward neural network with back-propagation training, applied for all the training and testing data, showing the accuracy of these type of classifiers, 96.6% overall. However, 98.39% of accuracy was obtained when using probabilistic neural works (NN-pnn), with 145 neurons in the radial basis layer. Can be also observed that the fuzzy inference and support vector machines methods do not achieve better results.

The drawback of the second step is that are unable to estimate the other components of the pollen. This second step can be used to confirm the result obtained in the first step, for the main component, or be used in a standalone configuration if is only needed to obtain the main pollen component.

The probabilistic neural networks provided the best results for the main Pollen classification. It is capable to obtain 98.39% of accuracy. When the secondary and tertiary pollen components are needed, the fuzzy inference system achieved better results. However, it is observed that the accuracy diminishes from the main component, 92.06%, to the tertiary component, 46.03%. For the pollen classification the secondary and tertiary pollen were added in order to include the predominant pollen and secondary pollen very important for the nutritional value of the final mixture of bee pollen.

To obtain a classification, given a physicochemical analysis of the pollen, as input, the (NN-pnn) takes 0.24 miliseconds. The (Fuzzy – Mamdani – FCM) method needs 2 miliseconds to obtain the classification. As such the better results in accuracy are also accompanied with the smallest processing time, for each sample data. The processing times were obtained with Matlab r2015b, running in a desktop computer, with i7 processor @ 3.60 GHz, and 64 Gb of RAM.

In the literature, do not exist methods to obtain the plant genus of bee pollen from its physicochemical properties. As such, a direct comparison is not possible. Similar studies do exist in the literature, e.g., for honey (Devillers, Morlot, Pham-Delègue, & Doré, 2004), which classify monofloral honeys based on their quality control data. Results obtained achieve 100% accuracy. The results observed in the work presented in this paper, for pollen, are in line with the honey application presented.

3.2. Physico chemical parameters estimation

With the aim of estimating the physicochemical parameters of the samples based on its previously ascertained plant genus (Table 3), it were successfully applied two computational methods. Following the previous sub-section methodology, fuzzy models and support vector machines obtained the best results, and are presented in Table 3. Neural networks were unable to estimate accurate models, i.e., R^2 indexes below 0.3 were obtained to estimate the physicochemical properties of bee pollen.

To estimate the above mentioned models, were successfully learned, a fuzzy mamdani inference with 100 clusters using FCM; and a support vector machine regression model using the linear epsilon insensitive cost, while the gaussian kernel's bandwidth has to be set to 0.1, along with the lagrangian multipliers bound set to 20. These parameters were obtained after a large set of experiments.

For the training and validation steps, the data was normalized and split with 80% for training and 20% for validation. A 5-fold cross validation approach was used, i.e., the data was randomly divided in training and validation, for five experiments, and the results are depicted in Table 3.

The results depicted in Table 3, are extremely relevant to solve the defined problem. In fact, the learned mamdani fuzzy inference system is capable to obtain the physicochemical characteristics of bee pollen, based only on the percentages of the plant genus. In other words, given a possible composition of plant genus, the learned model is capable to obtain the pollen physicochemical parameters with 95% confidence as presented by the F-test, for the R^2 and RMSE shown. This results suggest that the chemical composition of a mixture of pollen could be predict only with the predominant pollen (>45% of a specific pollen type).

For the results analysis were also obtained two more indexes, the Average I and Average II values that represent respectively, the mean of the physiochemical components related its composition (need for food product labeling) and the quality of the pollen, related to other characteristics that characterize the bee pollen quality but is not necessary for labeling. From these two indexes the most important for the honey producer is the second one, i.e., is more important to predict the quality of the pollen than its composition.

From Table 3, it is also observed that for the quality parameters, the model reached 0.85773 for R^2 with RMSE values lower than 10% when compared to the mean value of each parameter. This result is excellent for the complexity of the problem, i.e., 11 outputs and 11 inputs with only 210 data samples. Moreover, the observed prediction error gives an excellent insight for the producer about the quality of the pollen given the plant genus available for the bees to harvest in the hive surroundings.

This tool is very important and useful for the industry; because it is possible predict in an easy, fast and cheapest way the quality of a final mixture of bee pollen.

Table 3
Bee pollen physicochemical parameters estimation results given its plant genus.

Parameter	Mean ± σ	Min–max	CV	Fuzzy-Mamdani-FCM			SVM – regression		
				R^2	RMSE	F-test	R^2	RMSE	F-test
Moisture (%)	4.92 ± 0.95	2.76–7.24	19.36	0.7848	0.473	0	0.7554	0.449	1
pH	4.68 ± 0.62	3.40–5.98	13.21	0.6765	0.319	0	0.5693	0.445	1
Water activity	0.36 ± 0.08	0.23–0.53	21.19	0.7081	0.037	0	0.4846	0.059	1
Reducing sugars (%)	37.43 ± 6.40	22.43–54.21	17.09	0.866	2.416	0	0.4834	3.889	1
Proteins (%)	21.23 ± 3.54	12.76–29.68	16.69	0.6836	2.070	0	0.6232	2.370	0
Lipids (%)	5.00 ± 0.089	3.12–7.25	17.86	0.8535	0.371	0	0.6402	0.552	1
Ash (%)	2.6 ± 0.62	1.69–4.04	23.39	0.7996	0.297	0	0.5789	0.378	1
Fibre content (%)	3.24 ± 1.11	1.23–5.45	34.38	0.9276	0.279	0	0.7172	0.595	1
		Average I		0.78746			0.60653		
TPC (mg GAE/g pollen)	24.62 ± 6.52	12.57–44.96	26.49	0.8328	2.725	0	0.5433	2.936	0
Total flavonoids (mg QE/g pollen)	4.44 ± 1.63	2.05–10.78	36.73	0.8296	0.680	0	0.7931	2.885	1
EC_{50} (mg/g)	3.6 ± 1.33	0.60–6.71	36.94	0.9108	0.369	0	0.7953	0.755	1
		Average II		0.8577			0.7106		
		AVERAGE – all		0.8066			0.6349		

To obtain an estimation of physicochemical properties of organic compounds (Taskinen & Yliruusi, 2003), from observed data, is a difficult task. In this type of problems, RMSE, rarely surpasses 80% (Taskinen & Yliruusi, 2003) for the validation data set. As such, the results presented in this paper are in line with the ones presented in the literature, although these results are highly dependent on the application.

4. Conclusions

The paper successfully proposed computational methods based on NN, FCM and SVM, which allowed obtaining:

- The prediction of the physicochemical properties of bee pollen, by introducing some information regarding the plant genus composition of the samples. Results reached R^2 of 0.927 for some properties, and 0.857 for pollen quality;
- From the above result, the mamdani FM obtained allows estimating the quality of the analyzed bee pollen samples;
- A classification model that estimates predominant plant genus of a pollen, given its physicochemical properties, with accuracy of 98.39% obtained probabilistic NN;
- A mamdani fuzzy classification model to obtain up to the three principal genus of a pollen, given its physicochemical properties, with accuracy of 92.06%, 73.02%, 46.03% respectively.

The computational intelligence methods used have proven excellent results for the complexity of the problem, i.e., 11 outputs and 11 inputs with only 210 data samples. For the plant genus classification problem the systems have not shown adequate results for secondary classes of genus.

As future work, further pollen samples must to be gathered and analyzed to increase the accuracy and RMSE and R^2 values of the learned system, especially for the regression case. Since the system performs accurate predictions and classification, an application should be delivered to the bee pollen producers, to increase the quality of the produced pollen, and consequently for honey. More samples are required for the learned models, to improve its quality and robustness.

Acknowledgments

This work was partly supported by FCT, through IDMEC, under LAETA, project UID/EMS/50022/2013, and partly funded by FCT I.P.: Centro de Estudos Florestais, a research unit funded by FCT (UID/AGR/UI0239/2013); strategic programme UID/BIA/04050/2013 (POCI-01-0145-FEDER- 007569) and strategic programme UID/BIA/04050/2013 (POCI-01-0145-FEDER-007569). In addition, it was also funded by the ERDF through the COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI).

References

Amâcio, D., Serrano, M., Anjos, O., & Campos, M. (2014). Therapeutic potential of pollen. *Planta Medica, 80*(16).

Anjos, O., Iglesias, C., Peres, F., Martínez, J., García, Á., & Taboada, J. (2015). Neural networks applied to discriminate botanical origin of honeys. *Food Chemistry, 175*, 128–136.

AOAC (1995). *Official Methods of Analysis* (16th ed.). Arlington, VA, USA: Association of Official Analytical Chemists.

Arruda, V. A. S., Pereira, A. A. S., de Freitas, A. S., Barth, O. M., & de Almeida-Muradian, L. B. (2013). Dried bee pollen: B complex vitamins, physicochemical and botanical composition. *Journal of Food Composition and Analysis, 29*(2), 100–105.

Bárbara, M. S., Machado, C. S., Sodré, G. D. S., Dias, L. G., Estevinho, L. M., & de Carvalho, C. A. L. (2015). Microbiological assessment, nutritional characterization and phenolic compounds of bee pollen from *Mellipona mandacaia* smith, 1983. *Molecules, 20*(7), 12525–12544.

Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

Campos, M. G. R., Bogdanov, S., de Almeida-Muradian, L. B., Szczesna, T., Mancebo, Y., Frigerio, C., & Ferreira, F. (2008). Pollen composition and standardisation of analytical methods. *Journal of Apicultural Research, 47*(2), 154–161.

Campos, Maria Graça, Olena, Lokutova, & Anjos, O. (2016). Chemical Composition of bee pollen in chemistry, biology and potential applications of honeybee plant-derived products. In S. M. Cardoso & A. M. S. Silva (Eds.). Bentham Science Publishers.

Carpes, S. T., Begnini, R., de Alencar, S. M., & Masson, M. L. (2007). Study of preparations of bee pollen extracts, antioxidant and antibacterial activity. *Ciência E Agrotecnologia, 31*(6), 1818–1825.

Castilho, H. P., Goncalves, P. J. S., Pinto, J. R. C., & Serafim, A. L. (2007). Intelligent real-time fabric defect detection. image analysis and recognition. *Lecture Notes in Computer Science, 4633*, 1297–1307.

Chiu, S. L. (1994). Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems, 2*(3), 267–278.

Devillers, J., Morlot, M., Pham-Delègue, M. H., & Doré, J. C. (2004). Classification of monofloral honeys based on their quality control data. *Food Chemistry, 86*, 305–312.

Dhawale, V. R., Tidke, J. A., & Dudul, S. V. (2013). Neural network based classification of pollen grains. 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 79–84.

Estevinho, L. M., Rodrigues, S., Pereira, A. P., & Feás, X. (2012). Portuguese bee pollen: Palynological study, nutritional and microbiological evaluation. *International Journal of Food Science & Technology, 47*(2), 429–435.

Feas, X., Vazquez-Tato, M. P., Estevinho, L., Seijas, J. A., & Iglesias, A. (2012). Organic bee pollen: Botanical origin, nutritional value, bioactive compounds, antioxidant activity and microbiological quality. *Molecules, 17*, 8359–8377.

Feás, X., Vázquez-Tato, M. P., Estevinho, L., Seijas, J. A., & Iglesias, A. (2012). Organic bee pollen: Botanical origin, nutritional value, bioactive compounds, antioxidant activity and microbiological quality. *Molecules (Basel, Switzerland), 17*(7), 8359–8377.

Geisser, S. (1993). *Predictive inference*. New York, NY: Chapman and Hall.

Gonçalves, P. J. S. (2013). *The classification platform applied to mammographic images, computational intelligence and decision making*. Springer.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*(1–2), 273–324.

Li, P., & Flenley, J. R. (1999). Pollen texture identification using neural networks. *Grana, 38*, 59–64.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133.

Morais, M., Moreira, L., Feás, X., & Estevinho, L. M. (2011). Honeybee-collected pollen from five Portuguese Natural Parks: Palynological origin, phenolic content, antioxidant properties and antimicrobial activity. *Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association, 49*(5), 1096–1101.

Moreira, L., Dias, L. G., Pereira, J. A., & Estevinho, L. (2008). Antioxidant properties, total phenols and pollen analysis of propolis samples from Portugal. *Food and Chemical Toxicology, 46*(11).

Pascoal, A., Rodrigues, S., Teixeira, A., Feás, X., & Estevinho, L. M. (2014). Biological activities of commercial bee pollens: Antimicrobial, antimutagenic, antioxidant and anti-inflammatory. *Food and Chemical Toxicology: An International Journal Published for the British Industrial Biological Research Association, 63*, 233–239.

Pentoś, K., Łuczycka, D., & Wróbel, R. (2014). The identification of the relationship between chemical and electrical parameters of honeys using artificial neural networks. *Computers in Biology and Medicine, 53*, 244–249.

Ramzi, M., Kashaninejad, M., Salehi, F., Sadeghi Mahoonak, A. R., & Ali Razavi, S. M. (2015). Modeling of rheological behavior of honey using genetic algorithm-artificial neural network and adaptive neuro-fuzzy inference system. *Food Bioscience, 9*, 60–67.

Ranzi, A., Lauriola, P., Marletto, V., & Zinoni, F. (2003). Forecasting airborne pollen concentrations: Development of local models. *Aerobiologia, 19*, 39–45.

Salles, J., Cardinault, N., Patrac, V., Berry, A., Giraudet, C., Collin, M.-L., & Ellipsis Walrand, S. (2014). Bee pollen improves muscle protein and energy metabolism in malnourished old rats through interfering with the Mtor signaling pathway and mitochondrial activity. *Nutrients, 6*(12), 5500–5516.

Shafiee, S., Minaei, S., Moghaddam-Charkari, N., Ghasemi-Varnamkhasti, M., & Barzegar, M. (2013). Potential application of machine vision to honey characterization. *Trends in Food Science & Technology, 30*(2), 174–177.

Sousa, J. M. C., & Kaymak, U. (2002). *Fuzzy decision making in modeling and control*. Singapore: In World Scientific Co..

Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks, 3*, 109–118.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-15*(1), 116–132.

Taskinen, J., & Yliruusi, J. (2003). Prediction of physicochemical properties based on neural network modelling. *Advanced Drug Delivery Reviews*.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley-Interscience Publication.

Yang, K., Wu, D., Ye, X., Liu, D., Chen, J., & Sun, P. (2013). Characterization of chemical composition of bee pollen in China. *Journal of Agricultural and Food Chemistry, 61*(3), 708–718.

Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 30*(4), 451–462.