UV-VIS SPECTROSCOPY AND NEURAL NETWORKS IN DISTINGUISHING DIFFERENT TYPES OF HONEY

Kiril Nankov[1], Diana Tsankova[1], Georgi Terziyski[1]
[1]University of Food Technologies, Plovdiv, Bulgaria
e-mail: dtsankova@yahoo.com

Abstract: The aim of the article is to investigate the potential of honey discrimination (on the base of its botanical origins) by UV-Vis spectroscopy and neural networks. One hundred and eleven samples from three types of honey (acacia, linden, and honeydew) are measured by a spectrophotometer "Cary100" with recorded wavelength range of 190~900 nm. The samples are used as input data for two types of honey classifiers: the one is based on linear discriminant analysis (LDA), and the other uses an artificial neural network. The neural classifier is realized as a multilayered perceptron with Backpropagation learning algorithm. Principal components analysis (PCA) is used for reducing the number of inputs and for a proper visualization of the experimental results. The comparative analysis of the proposed classifiers is based on leave-one-out-cross validation test.
Key words: UV-Vis spectroscopy, honey discrimination, PCA, LDA, artificial neural network

INTRODUCTION

"Honey is the natural sweet substance, produced by honeybees from the nectar of flowers or from secretions of living parts of plants or excretions of plant sucking insects on the living parts of plants, which the bees collect, transform by combining with specific substances of their own, deposit, dehydrate, store and leave in honeycombs to ripen and mature" [1, 2]. Honey consists of sugars, water, amino acids, oil, mineral salts and especial enzymes produced by bees [3].
It is possible to determine the origin of specific samples of honey and the environmental pollution of a region from the quantitative and qualitative ratio of heavy and rare metals in honey [4]. Honey is proven to be a highly sensitive bioindicator for estimation of the environmental pollution with SO_2 [5]. It is possible to assume that honeybees (Apis mellifera L.) respond to changes in their environment and in particular to increased quantities of heavy metals in soil, air, plants. That makes them a reliable indicator and allows their use in biomonitoring of the environment [6].
Bee pollen appears to be safe for most people, but if someone has pollen allergies, bee pollen can cause a serious allergic reaction. Bee pollen is not safe for pregnant women and also may cause increased bleeding if taken with certain blood thinners [7]. Proteins derived from saliva of honeybee and pollen proteins contained in the honey cause allergic reactions to honey [8]. Different types of pollen (defining different floral origin of honey) can cause different allergic reactions in different people. Therefore it should be developed a means of distinguishing between different types of honey.
Several methods have been used for the determination of the floral origin of honey and among them the pollen recognition and sensory analysis are the most popular ones. However the technique of analysis of honey's pollen content is tedious and has some limitations. The other methods are mainly based on the analysis of honey's aroma compounds, sugar profile, flavonoid pattern, non-flavonoid phenolics, organic acids, isotopic relations, and protein and amino acid compositions and marker presence [9, 10]. But some of these methods are generally too time-consuming, complex, and labour intensive for routine quality control application or require very specialized personnel to interpret the results [11, 12].
In addition, most of the analytical techniques involve some kind of sample pre-treatment. The advantages of the technique of UV-visible (UV-Vis), infrared (IR) and fluorescence spectroscopy with respect to other analytical methods are the non-invasive approach, the relatively easy and quick data acquisition. Some authors [13, 14] have used the IR technique for qualification of adulterants in honey with good accuracy. Recently, both near infrared (NIR) and middle infrared (MIR) spectroscopy, were successfully used for classification of unifloral and multifloral honeys [15, 16, 17, 18]. Some authors have used Vis spectrometry for the same purpose [19, 20, 21], but there is almost no information on the use of UV for classification of honey according to its botanical origin [22].

The purpose of the paper is to investigate the possibility of honey discrimination (based on its botanical origin) using UV-Vis spectroscopy in absorbance mode. Spectroscopic data obtained undergo subsequent statistical processing including principal components analysis (PCA) for reducing the classifiers' number of inputs. An artificial neural network (ANN) with Backpropagation (BP) learning algorithm is proposed to classify honey in appropriate classes related to its floral origin. Another type of classifier is considered, which uses the popular linear discriminant analysis (LDA) to provide a basis for comparison of the results obtained. The performance of the two calibration models is confirmed by leave-one-out-cross validation test in MATLAB environment.

MATERIAL AND METHODS

Honey spectrum acquisition

One hundred and eleven samples of three different types of honey (acacia – 34 samples; linden – 46 samples; and honeydew – 31 samples) were purchased from supermarkets and from private producers. All samples of honey were diluted with distilled deionized water till 10% solution. The samples were annealed at room temperature (23-24^{0}C). The spectral characteristics of the honey were taken with a spectrophotometer Cary100 ranging from 190 to 900 nm at 1 nm sampling space.
Spectral readings of the three types of honey were treated by the PCA to reduce dimensionality of the input data, then they were classified using two classifiers – LDA and ANN based ones. In addition, the influence of the individual spectrum ranges - UV only, Vis only (plus very near infrared, i.e. Vis+), and the merging of both areas (UV and Vis +) on the classification result was investigated.

Principal Components Analysis [23, 24]

The aim of the method is to reduce the dimensionality of multivariate data (e.g., wavelengths) whilst preserving as much of the relevant information as possible. PCA is a linear transformation that transforms the data (observations of possibly correlated variables) to a new coordinate system such that the new set of variables, the principal components, are linear functions of the original variables. Principal components (PCs) are uncorrelated, and the greatest variance by any projection of the data comes to lie on the first coordinate, the second greatest variance on the second coordinate, and so on. This is achieved by computing the covariance matrix for the full data set. Then, the eigenvectors and eigenvalues of the covariance matrix are computed, and sorted according to decreasing eigenvalue [23, 24]. All the principal components are orthogonal to each other. The full set of principal components is as large as the original set of variables. Usually the sum of the variances of the first few principal components exceeds 80% of the total variance of the original data [25].

Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a classic classifier, with, as its name suggests, a linear decision surface. The basic idea of LDA is to find a linear transformation, such that the ratio of the between-class scatter and the within-class scatter is maximized. Samples are projected to a new space with smallest within-class distance and largest inter-class distance [26]. Although LDA usually gives a good discrimination performance, it suffers from some deficiencies if variables are highly correlated or class boundaries are complex or nonlinear. To avoid such deficiencies, in the former case, variables are often transformed by correlation-reducing methods such as PCA, and in the latter case, LDA could be replaced by quadratic discriminant analysis or ANN.

Artificial neural network based classifier

It is well known that artificial neural networks with a feedforward multilayered structure are universal function approximators [27, 28]. One classification task can be easily reduced to a task for approximation. Let the classifier of honey be implemented as a neural network with a feedforward structure and Backpropagation (BP) learning algorithm. The neural network consists of $n = 2$ inputs (n is the number of PCs used by classifiers), 3 outputs and 2 hidden layers. The three outputs of the

network correspond to the three classes of honey: acacia, linden and honeydew honey. The two hidden layers contain neurons with 'tansigmoid' activation function (hyperbolic tangent), and the activation function of the three output neurons is 'logsigmoid' [29].
The method of the PCs is applied to UV and Vis+ spectra and total n = 2 PCs are used as input training samples for the neural network. The supervisor supplies the network's output with the following three combinations: '1 0 0', '0 1 0' or '0 0 1', depending on whether the input receives the data for the classes 'acacia', 'linden' or 'honeydew', respectively. BP is a gradient-based learning algorithm that minimizes the sum squared error between the real and required input of the ANN.

RESULTS AND DISCUSSION

Absorbance Spectra

After receiving the "raw" data from the measurements by spectroscopy, they are subject to pre-treatment. In order to remove some apparent interference received data are limited from above by a predetermined value. The resulting absorbance curves are smoothed by the method of creeping averaging, using the formula:

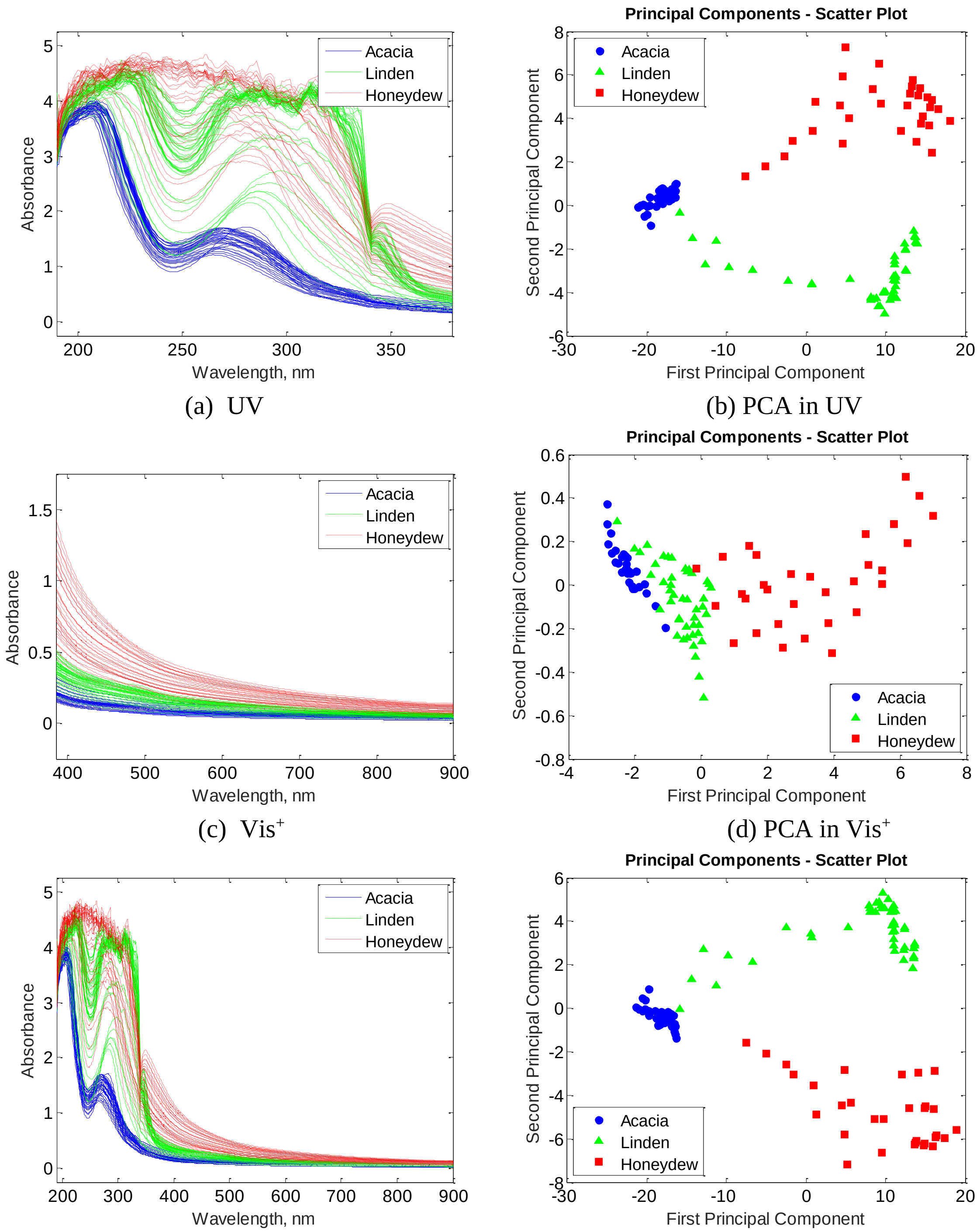
$$a_{i+l/2} = \frac{1}{l+1}\sum_{k=0}^{l} a_{i+k}\ , \tag{1}$$

where l is the width of a linear filter accepting even-numbered values.
In the experiments shown below, three absorption spectra were used: (1) UV from 190 to 380 nm, (2) Vis$^+$ from 380 to 900 nm, and (3) UV-Vis$^+$ from 190 to 900 nm (Fig.1a,c,e). The PCA was applied to the spectral characteristics of the three ranges. The spectral dimensionality was reduced to a small number (two) of PCs using PCA. The scores scatter plots of the 1st and 2nd PCs are shown in Fig.1b,d,f, where samples from classes 'acacia', 'linden' and 'honeydew' are marked with circular, triangular and squared symbols, respectively. It is evident that the samples form three clusters, that with a few exceptions coincide with the three types of honey mentioned above. Here, determining the type of honey is based solely on the inscription on the label by the manufacturer, i.e. trusting the manufacturer. The first two PCs explain as high as: (a) 98.15 % of variance of the spectra (93.10 % for PC-1 and 5.06 % for PC-2) for UV region; (b) 99.89 % of variance of the spectra (99.41 % for PC-1 and 0.48 % for PC-2) for Vis$^+$ region; and (c) 97.89 % of variance of the spectra (91.83 % for PC-1 and 6.07 % for PC-2) for UV-Vis$^+$ one. The two PCs were chosen to develop the classification models.

LDA and ANN Based Honey Classifiers

On the basis of these two PCs honey a few classifiers were developed using LDA and ANN. The prediction results of the honey's botanical origin made by the proposed classifiers, UV-LDA and UV-ANN, are shown in Fig. 2 and Table 1. The performance of the UV-ANN based model is better (96.4 % accuracy) than the UV-LDA based one (91.9 %) for honey discrimination. In the case of the better model (Table 1), 1 sample from observed class 'linden' was predicted wrong as 'acacia', while 3 samples from the same class 'linden' were predicted wrong as 'honeydew'. The model predicted 107 out of 111 samples correctly. The prediction accuracy 96.4 % was divided between the classes as follows: 91.3 % for class 'linden', and 100 % for the other two classes - 'acacia' and 'honeydew'.
The prediction results of the type of honey made by the Vis$^+$-LDA and Vis$^+$-ANN classifiers are shown in Fig. 3 and Table 2. Again, the ANN based classifier had better performance than the LDA based one - 93.7 % vs. 83.8 %. But the success rate was less than the result in UV domain. The Vis$^+$-ANN based classifier predicted wrong 3 samples from observed class 'acacia' as 'linden', 2 samples from class 'linden' as 'acacia', and 2 'honeydew' samples – as 'linden'. The model predicted 104 out of 111 samples correctly. The prediction accuracy of each of the three classes was: 91.2 % 'acacia', 95.7 % 'linden', and 93.5 % 'honeydew'.
After the combination of the two spectral regions UV and Vis$^+$ (Fig.4 and Table 3), the following prediction results were obtained: 99.1% accuracy for UV-Vis$^+$-ANN based classifier and 93.7% for UV-Vis$^+$-LDA based one. The ANN based model predicted 110 out of 111 samples correctly. Only 1

sample from observed class 'linden' was predicted wrong as 'acacia'. Thus the prediction accuracy of the classes 'acacia' and 'honeydew' was 100 %, while that of the class 'linden' - 97.8 %.
Table 4 summarizes the results of all honey classifications on the basis of botanical origin considered above. As seen the highest prediction results were obtained by using both UV and Vis+ regions in the classifier based on an artificial neural network.

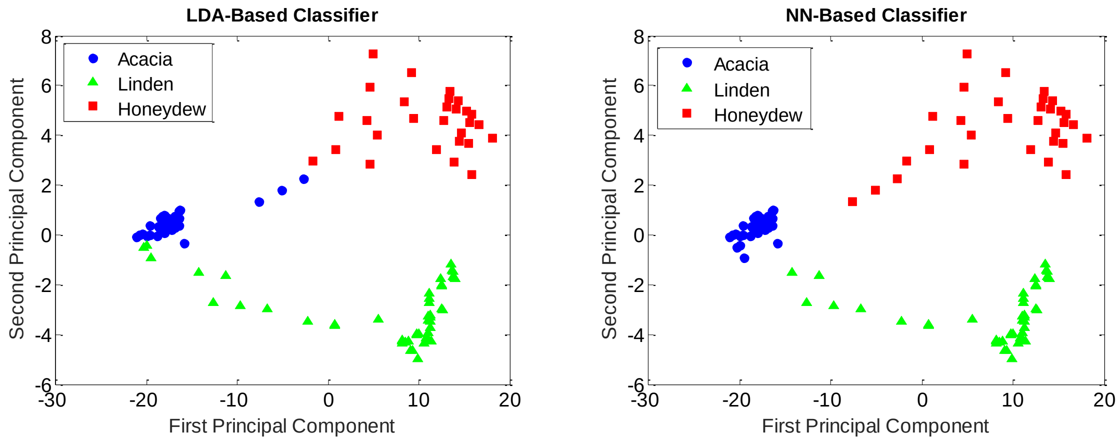


(a) UV (b) PCA in UV

(c) Vis$^+$ (d) PCA in Vis$^+$

(e) UV-Vis$^+$ (f) PCA in UV-Vis$^+$

Figure 1. Absorbance spectra of honey in UV, Vis$^+$, UV-Vis$^+$ domains, and corresponding PCA scatter plots

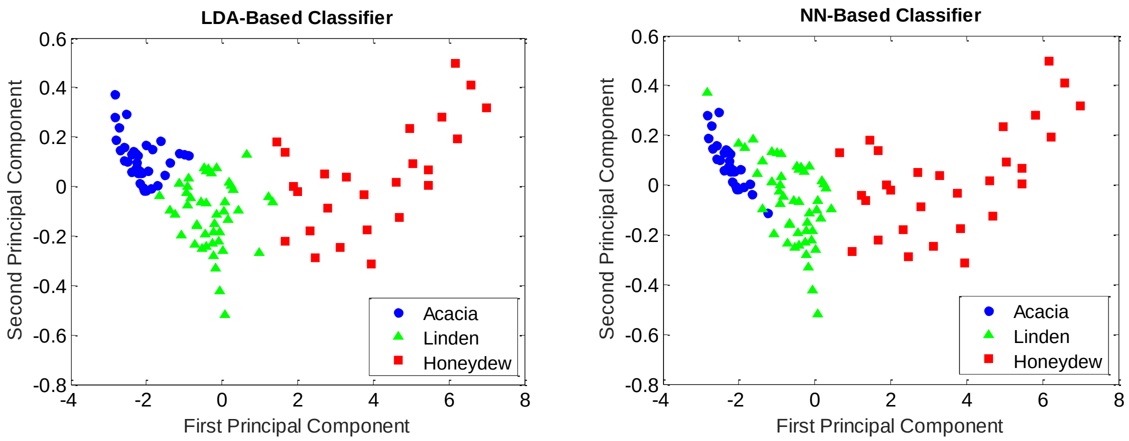


(a) (b)

Figure 2. UV-LDA and UV-ANN Based Honey Classifiers: (a) LDA-based classifier; (b) ANN-based one

Table 1. Discrimination accuracy of UV-LDA and UV-ANN based models

		Predicted Class by UV-LDA				Predicted Class by UV-ANN		
		Acacia	Linden	Honeydew	←→	Acacia	Linden	Honeydew
Observed Class	Acacia	34	0	0	34	34	0	0
	Linden	4	39	3	46	1	42	3
	Honeydew	2	0	29	31	0	0	31
		40	39	32	111	35	42	34
		Success = 91.8919 %				Success = 96.3964 %		

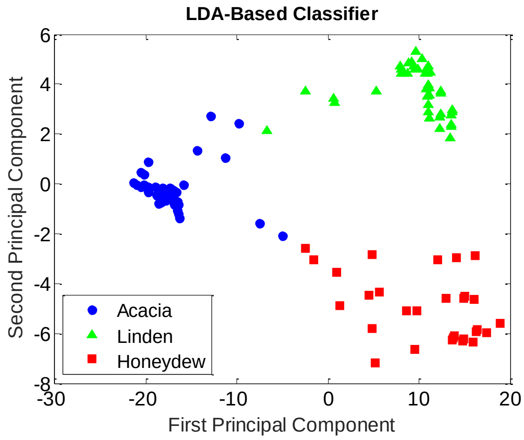


(a) (b)

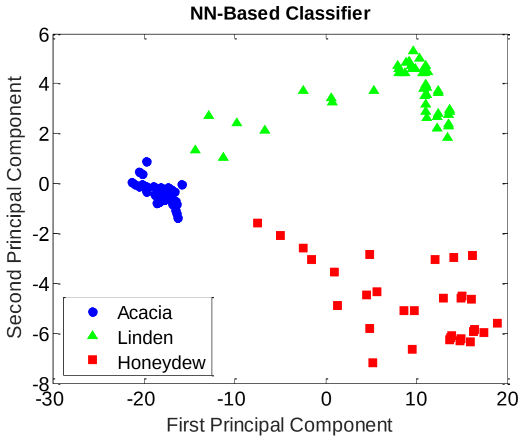
Figure 3. Vis$^+$-LDA and Vis$^+$-ANN Based Honey Classifiers: (a) LDA-based classifier; (b) ANN-based one

Table 2. Discrimination accuracy of Vis$^+$-LDA and Vis$^+$-ANN based models

		Predicted Class by Vis$^+$-LDA				Predicted Class by Vis$^+$-ANN		
		Acacia	Linden	Honeydew	←→	Acacia	Linden	Honeydew
Observed Class	Acacia	31	3	0	34	31	3	0
	Linden	9	37	0	46	2	44	0
	Honeydew	0	6	25	31	0	2	29
		40	46	25	111	33	49	29
		Success = 83.7838 %				Success = 93.6937 %		



(a)



(b)

Figure 4. UV-Vis$^+$-LDA and UV-Vis$^+$-ANN Based Honey Classifiers: (a) LDA-based classifier; (b) ANN-based one

Table 3. Discrimination accuracy of UV-Vis$^+$-LDA and UV-Vis$^+$-ANN based models

		Predicted Class by UV-Vis$^+$-LDA				Predicted Class by UV-Vis$^+$-ANN		
		Acacia	Linden	Honeydew	←→	Acacia	Linden	Honeydew
Observed Class	Acacia	34	0	0	34	34	0	0
	Linden	5	41	0	46	1	45	0
	Honeydew	2	0	29	31	0	0	31
		41	41	29	111	35	45	31
		Success = 93.6937 %				Success = 99.0991 %		

Table 4. Comparative analysis of LDA and ANN-based classifiers for all the ranges

No.	Spectral range	LDA-based classifier	ANN-based classifier
1	UV	91.9 %	96.4 %
2	Vis$^+$	83.8 %	93.7 %
3	UV-Vis+	93.7 %	99.1 %

CONCLUSION

In this article the spectral properties of Bulgarian honey were investigated in regard to the potential of honey discrimination on the base of its botanical origin. The UV and Vis spectra separately and together were used for training two honey classifiers – LDA based classifier and ANN based one. By means of PCA the number of input data was reduced to only 2 PCs. The obtained advantages are the lack of correlation between the input data and also the ability to visualize the clusters formed by different types of honey. The ANN based classifier using wavelengths ranging from 190 to 900 nm (UV-Vis+ spectra) shows the best prediction accuracy, 99.1%, determined by the 111 leave-one-out-cross-validation tests.
Future work will include methods for distinguishing honey not only on the basis of its botanical origin but also on the basis of its geographical origin. Honey will be analyzed in areas around coal-fired power plants and fields contaminated with heavy metals.

REFERENCES

[1] Codex Alimentarius: Draft revised standard for honey (at step 10 of the Codex procedure). Alinorm 01/25, pp.19-26, 2001.

[2] EU Council: Council directive 2001/110/EC of 20 December 2001 relating to honey. Official Journal of the European Communities L10, pp.47–52, 2002.
[3] Boffo, E.F., L.A. Tavares, A.C.T. Tobias, M.M.C. Ferreira, A.G. Ferreira. Identification of components of Brazilian honey by 1H NMR and classification of its botanical origin by chemometric methods, LWT - Food Science and Technology, Vol.49, pp.55-63, 2012, www.elsevier.com/locate/lwt
[4] Birutė, S., M. Paulius, B. Rūta, Honey as an Indicator of Environmental Pollution. Environmental Research, Engineering & Management, Vol.36, No.2, pp.53-58, 2006.
[5] Ponikvar-Svet, M., J. Šnajder, B. Sedej, Honey as a bioindicator for environmental pollution with SO_2, Apidologie, Vol.36, No.3, 2005, DOI: 10.1051/apido:2005027
[6] Zhelyazkova, I., Honeybees – bioindicators for environmental quality, Bulgarian Journal of Agricultural Science, Vol.18, No 3, pp.435-442, 2012, Agricultural Academy, Bulgaria.
[7] Ratini, M., Bee Pollen Benefits and Side Effects, WebMD Medical Reference Reviewed by Melinda Ratini, DO, MS on January 26, 2018, https://www.webmd.com/balance/bee-pollen-benefits-and-side-effects
[8] Bauer, L., A. Kohlich, R. Hirschwehr, U. Siemann, H. Ebner, O. Scheiner, D. Kraft, C. Ebner, Food allergy to honey: Pollen or bee products? Characterization of allergenic proteins in honey by means of immunoblotting, The Journal of Allergy and Clinical Immunology, Vol. 97, No.1, pp. 65–73, 1996.
[9] Anklam E., A review of the analytical methods to determine the geographical and botanical origin of honey. Food Chem., Vol.63, pp.549-562, 1998.
[10] Hermosin, I., R.M. Chicon, M.D. Cabezudo. Free amino acid composition and botanical origin of honey, Food Chemistry, Vol.83, pp.263–268, 2003, www.elsevier.com/locate/foodchem
[11] Bogdanov S., P. Martin. Honey authenticity, Mitt. Geb. Lebensmittelunters Hyg., Vol.93, pp. 232–254, 2002.
[12] Benedetti, S., S. Mannino, A.G. Sabatini, G.L. Marcazzan. Electronic nose and neural network use for the classification of honey, Apidologie, Vol.35, pp.1–6, 2004, INRA/DIB-AGIB/ EDP Sciences, DOI: 10.1051/apido:2004025
[13] Gallardo-Velazquez, T., G. Osorio-Revilla, M.Z.D. Loa, Y. Rivera-Espinoza. Application of FTIR-HATR spectroscopy and multivariate analysis to the quantification of adulterants in Mexican honeys, Food Research International, Vol.42, No.3, pp.313–318, 2009.
[14] Zhu, X., S.Li, Y. Shan et al. Detection of adulterants such as sweeteners materials in honey using near-infrared spectroscopy and chemometrics, Journal of Food Engineering, Vol.101, No.1, pp. 92–97, 2010.
[15] Davies A., B. Radovic, T. Fearn, E. Anklam. A preliminary study on the characterisation of honey by near infrared spectroscopy, J. Near Infrared Spectrosc., Vol.10, pp.121–135, 2002.
[16] Ruoff K., W. Luginbühl, S. Bogdanov, B. Estermann, T. Ziolko, R. Amado. Potential of Near Infrared Spectroscopy for Authenticity Testing of Unifloral Honey, Eur. Congr. for Authenticity of Food, Nyon, Switzerland, 2003.
[17] Lichtenberg-Kraag B. Infrared spectroscopy: The quality assurance in honey analysis, Apidologie, Vol.34, pp. 479–480, 2003.
[18] Lekova, S., D. Tsankova, Determination of Botanical Origin of Honey by Mid Infrared Spectroscopy (Mid-FTIR), Colorimetry and Chemometric Analysis, Journal of Chemical Technology and Metallurgy (JCTM), Vol. 52, No. 1, pp. 52 - 57, 2017, University of Chemical Technology and Metallurgy, BG, http://dl.uctm.edu/journal/node/79
[19] Li, Y., H. Yang. Honey Discrimination Using Visible and Near-Infrared Spectroscopy, International Scholarly Research Network, ISRN Spectroscopy, Article ID 487040, 4 pages, 2012, doi:10.5402/2012/487040
[20] Tsankova, D., S. Lekova, Botanical Origin-Based Honey Discrimination Using Vis-NIR Spectroscopy and Statistical Cluster Analysis, Journal of Chemical Technology and Metallurgy (JCTM), Vol. 50, No. 5, pp. 638-642, 2015, University of Chemical Technology and Metallurgy, BG, http://dl.uctm.edu/journal/node/71

[21] Tsankova, D., S. Lekova, K. Nikolova, G. Terziyski, VIS Spectroscopy-Based Chemometric Analysis of Honey with Respect to Discrimination of Its Botanical Origin, Annals of Faculty Engineering Hunedoara – International Journal of Engineering, Romania, 2015, Vol. XIII, No.4, pp. 275-279, ISSN: 1584-2673 (CD edition, online).

[22] Tsankova, D., S. Lekova, UV-Vis Spectroscopy and Statistical Cluster Analysis in Distinguishing Different Types of Honey, International Conference "Automatics and Informatics'2015", Sofia, Bulgaria, 4 - 7 October 2015, pp.107-110.

[23] Hotelling, H., Analysis of a Complex of Statistical Variables into Principal Components. Journal of Educational Psychology, Vol.24, No.6 & 7, pp. 417–441 & pp. 498–520, 1933.

[24] Jolliffe, I. T., Principal Component Analysis. Second ed. Springer Series in Statistics. New York: Springer-Verlag New York, 2002.

[25] Statistics Toolbox™ User's Guide, R2014a, © COPYRIGHT 1993–2014 by The MathWorks, Inc.

[26] Kim, H., Drake, B. L., Park, H., Multiclass classifiers based on dimension reduction with generalized LDA, Pattern Recognition, Vol. 40, No. 11, pp. 2939–2945, 2007.

[27] Blum, E.K., Li, L.K., Approximation theory and feedforward networks, Neural Networks, Vol.4, pp.511-515, 1991.

[28] Hornik, K., Stinchcombe, M., White, H., Multilayer feedforward networks are universal approximators, Neural Networks, Vol.2, pp.359-366, 1989.

[29] Beale, M.H., Hagan, M.T., Demuth, H.B., Neural Network ToolboxTM, User's Guide, The MathWorks Inc., 1992-2015. https://www.mathworks.com/help/pdf_doc/nnet/nnet_ug.pdf