Grana

ISSN: 0017-3134 (Print) 1651-2049 (Online) Journal homepage: http://www.tandfonline.com/loi/sgra20

An automatic identification method for the comparison of plant and honey pollen based on GLCM texture features and artificial neural network

Y. Kaya , M. E. Erez , O. Karabacak , L. Kayci & M. Fidan

To cite this article: Y. Kaya , M. E. Erez , O. Karabacak , L. Kayci & M. Fidan (2013) An automatic identification method for the comparison of plant and honey pollen based on GLCM texture features and artificial neural network, Grana, 52:1, 71-77, DOI: 10.1080/00173134.2012.754050

To link to this article: https://doi.org/10.1080/00173134.2012.754050

Published online: 12 Feb 2013.

Grana, 2013
Vol. 52, No. 1, 71–77, http://dx.doi.org/10.1080/00173134.2012.754050

An automatic identification method for the comparison of plant and honey pollen based on GLCM texture features and artificial neural network

YILMAZ KAYA[1], MEHMET EMRE EREZ[2], OSMAN KARABACAK[2], LOKMAN KAYCI[2] & MEHMET FIDAN[2]

[1] *Siirt University, Engineering and Architecture Faculty, Computer Engineering, Siirt, Turkey,* [2] *Siirt University, Faculty of Science and Art, Department of Biology, Siirt, Turkey*

Abstract
Pollen grains vary in colour and shape and can be detected in honey used as a way of identifying nectar sources. Accurate differentiation between pollen grains record is hampered by the combination of poor taxonomic resolution in pollen identification and the high species diversity of many families. Pollen identification determines the origin and the quality of the honey product, but this indefiniteness is also a big challenge for the beekeepers. This study aimed to develop effective, accurate, rapid and non-destructive analysis methods for pollen classification in honey. Ten different pollen grains of plant species were used for the estimation. GLCM (grey level co-occurrence matrix) texture features and ANN (artificial neural network) were used for the identification of pollen grains in honey by the reference of plant species pollen. GLCM has been calculated in four different angles and offsets for the pollen of the plant and the honey samples. Each angle and offset pair includes five features. At the final step, features were classified using the ANN method; the success of estimation with ANN was 88.00%. These findings suggest that the texture parameters can be useful in identification of the pollen types in honey products.

Keywords: *Honey, pollen identification, expert system, GLCM, artificial neural network*

Honey is produced by honeybees from nectar of plants and contains at least 181 substances; it is considered as an important part of traditional medicine (Louveaux, 1985; Sato & Miyata, 2000; Ferreira et al., 2009). The growing interest in honey associated with the variety of identifications has been observed in the domain of nutrition science for several years. Pollen analysis appears to be the most frequently used method in identification of honey region studies (Bolchi Serini & Salvi, 1990; Bambara, 1991; Szczęsna & Rybak-Chmielewska, 1993).

Pollen characterisation and quantification of honeys have been the traditional methods to determine the nectar sources and botanical origin, but this technique is time-consuming and has some limitations (von der Ohe, 1994; Hermosin et al., 2003). Although the palynological characterisation method is important for classifying the honey flavour characteristics, it is time-consuming, requires a very experienced analyst and is strongly dependent on the expert's ability and judgment (Piana et al., 2004). The resistant outer microspore wall of pollen grains has various three-dimensional shapes, which allow identifying the plant species. Pollen in honey can be used to determine the origin and the quality of the honey product. Here, we use a consistent methodology for the identification of honey pollen through comparison of plant species pollen by using an expert system.

A number of attempts have been made to develop computer tools to help the classification of pollen types. Many past studies were little more than automated classification processes. While automated

Correspondence: Yılmaz Kaya, Siirt University, Engineering and Architecture Faculty, Computer Engineering, 56100, Siirt, Turkey. E-mail: *yilmazkaya 1977@gmail.com*

(Received 5 April 2012; accepted 22 September 2012)

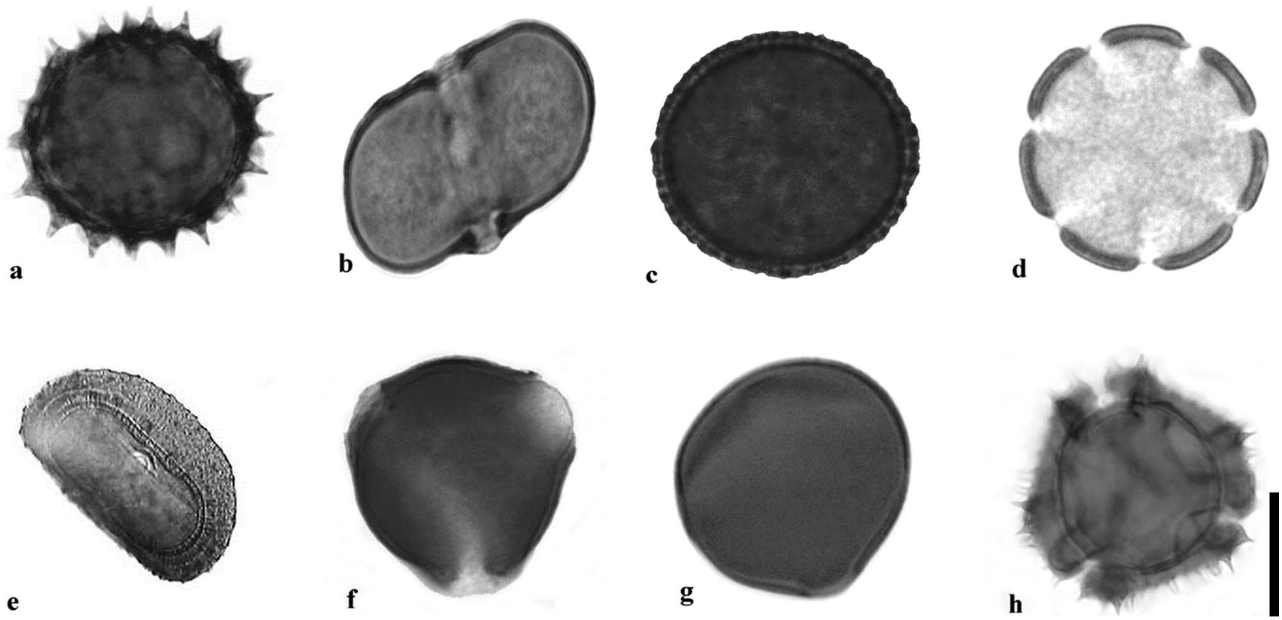

Figure 1. Examples of selected pollen grains from plants and honey. **A.** *Gundelia tournefortii* L. **B.** *Anchusa azurea* R. Mill. **C.** *Daphnea mucronata* Royle. **D.** *Asperula xylorrhiza* Nab. **E.** *Echinops pungens* Trautv. **F.** *Rosa canina* L. **G.** *Astragalus brachycalyx* Fischer. **H.** *Cichorium intybus* L. Scale bar – 2 μm.

classifications were the first step in automating identification of pollen grains, new techniques such as artificial intelligence systems may provide better solutions in the long term. Several research studies about pollen recognition have been published using scanning electron microscopy (SEM) and some good results have been achieved (Vezey & Skvarla, 1990; France et al., 2000), but SEM analysis is expensive, slow and absolutely not suitable for a real-time application. Langford et al. (1990) developed a computer system whose purpose was the classification of pollen grains based on analysis of their surface texture. The system was based on the use of scanning electron microscope images of grain surfaces. Li and Flenley (1999) have described initial work on texture analysis using light microscope images. In a study, they tried to differentiate aerobiological spores by image analysis (Benyon et al., 2000). Recently, works have been presented on pollen recognition using two-dimensional (2D) statistical classification (Jones, 2000) or using three-dimensional (3D) grey scale invariants with confocal microscopy (Ronneberger, 2000).

Texture analyses of biological images are computer-based techniques that extract features using a mathematical approach to the distribution grey level of pixels in images. The textures of biological images from different plant species may vary from each other. Grey-level images carry important information belonging to the textures of the images. Grey level co-occurrence matrix (GLCM) is a feature extraction method that allows measurement of some statistical characters of an image (Guo et al., 2008). GLCM is used in many different areas such as medicine (Xian, 2010), engineering, biology, agriculture (Huang, 2007) or remote sensing (Guo et al., 2008). Artificial neural network (ANN) systems are computer systems that automatically develop skills, without demanding any help, such as composing and discovering new data with a training way which is one of the features of the human brain (Kaya et al., 2010).

Material and methods

Material and pollen preparation

In this study, plant pollen from 50 species was collected from wild populations in Pervari, Turkey. Plants were collected in the flowering season and identified by specimens based on the 'Flora of Turkey and the East Aegean Islands' (Davis, 1970). Pollen slides were prepared using the technique of Wodehouse (1935). The pollen was first treated with 70% alcohol to remove oily substances, and then embedded in glycerine jelly stained with basic fuchsine. This material was used to prepare slides by using glycerine-jelly mixed with 1% Safranin (Erdtman, 1969).

Honey samples were taken from hives directly and stored at the laboratory at 4 °C. One gram of honey was dissolved in distilled water and heated to just below 50 °C. Dissolved samples were centrifuged at 200*g* for 20 minutes. Supernatants were poured off and the residues were mixed with safranine glycerine medium and examined using light microscopy. Pollen from honey and plants were compared with each other to decide which pollen would be suitable for the expert system (Figure 1). Selection of pollen species was determined according to the most abundant pollen in honey. Consequently, ten pollen types were considered to study. The prepared slides of honey and selected plant pollen were studied under an Olympus CX31 light microscope using oil immersion. Different positions of pollen images were photographed and image definitions were adjusted with the Stream Start software package.

Proposed method for pollen identification

A method based on level co-occurrence matrix (GLCM) and ANN is used for the identification of pollen grains in honey by the reference of plant species pollen. One hundred pollen images were used from ten different species. Fifty pollen images from plants (training set) and fifty pollen images from honey (test set) were obtained. Each of the pollen images was cropped to 512×512 pixels before processing. The texture features were used for the classification of pollen grains. Five texture features (homogeneity, contrast, energy, correlation and entropy) of the images were used with ANN for identification and classification of pollen grains. ANN uses a biological neural system for modelling. The study consists of five consecutive blocks (Figure 2). The processes in these blocks can be briefly summarised as follows:

- *Block 1:* Obtaining of five images belonging to the different positions of ten selected pollen grains for plant species and honey.
- *Block 2:* Obtaining contrast, correlation, entropy, energy and homogeneity of pollen textures from GLCM.
- *Block 3:* Fifty pollen images from plants were considered as training set and fifty pollen images from honey were considered as test set. The aim of these sets is to compare pollen from honey with pollen of plant species.
- *Block 4:* Classification of training-test partitions through ANN.
- *Block 5:* Presentation of classification results (decision stage).

Extraction of texture features of pollen grains from GLCM

Analyses of texture features have been widely employed in classification, proper feature selections, classifier design can be greatly simplified. Hence, we adopt texture features from the GLCM (Haralick

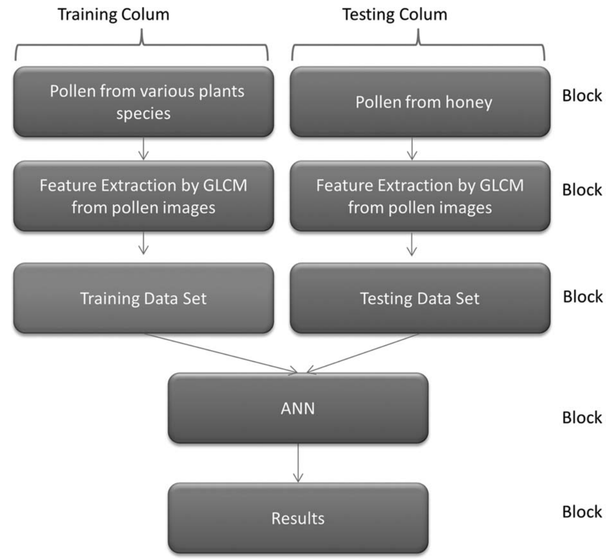

Figure 2. Proposed method for pollen identification of honey and plant samples by GLCM and ANN techniques.

et al., 1973) to identify pollen grains. GLCM is a pixel-based image processing method. An element $p(\varphi, d, i, j)$ of a GLCM of an image represents the relative frequency, where i is the grey level at location (x, y), and j represents the grey level of a neighbouring pixel at a distance d and an orientation φ from location (x, y). The creation of the GLCM matrix is based on the distance between pixels, the pixels angle (0°, 45°, 90° and 135°) and the number of level grey-scale conversion done (maximum 256) parameters (Ondimu & Murase, 2008). In this study, homogeneity, contrast, energy, correlation and entropy were the statistical texture features computed using GLCMs. GLCM depending angle and distance parameters is expressed in Equation (1) (Haralick et al., 1973).

$$\begin{aligned} P(i,j,d,\varphi^o) &= \#\{((k,l),(m,n)) \in D, \\ &((k-m),(l-n)) \in \{-d,0,d\}, I(k,l) = i, \\ &I(m,n) = j, \angle((k,l),(m,n)) = \varphi^o\} \end{aligned} \tag{1}$$

The GLCMs of distance ($d = 1$, 2, 3, 4) pixel and orientations 0°, 45°, 90°, and 135° are used for subimage, with resolution 512×512 pixels. The texture features used in this study, are computed as follows:

(I) *Contrast.* – Contrast is a measurement of the local variations present in the image.

$$f_1 = \sum_{m=0}^{N_g-1} m^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p_{d,\vartheta^o}(i,j) \right\}, \quad |i-j| = m \tag{2}$$

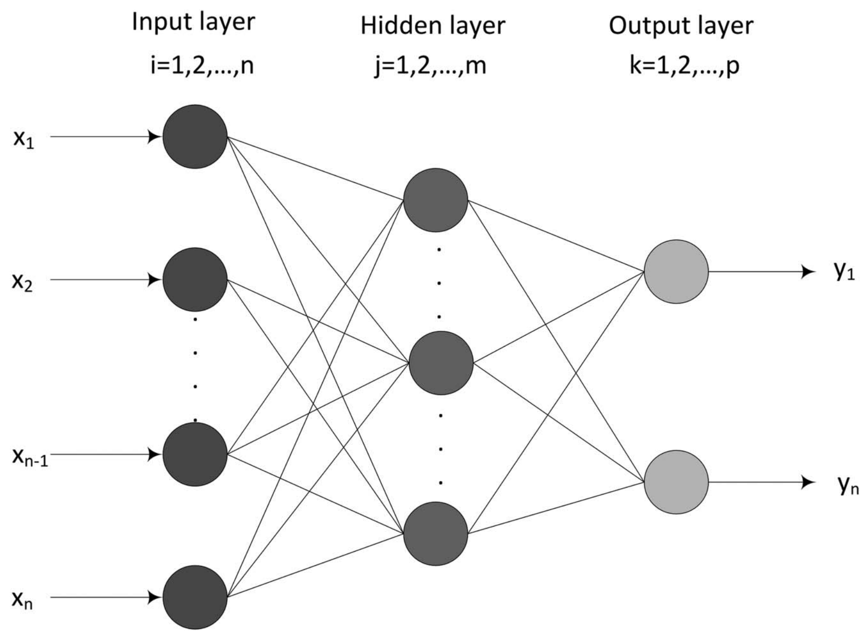

Figure 3. The basic architecture of the applied ANN method for pollen identification.

(II) *Correlation.* – Correlation is an indication of the linearity of the relationship of grey levels of pixel pairs.

$$f_2 = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(ij)p_{d,\vartheta^o}(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{3}$$

(III) *Entropy.* – Entropy measures disorder of the image and it indicates complexity within an image. When the image is not texturally uniform, entropy is very large.

$$f_3 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_{d,\vartheta^o}(i,j)\log p_{d,\vartheta^o}(i,j) \tag{4}$$

(IV) *Energy.* – Energy expresses the repetition of pixel pairs of an image. It is a measure of homogeneity of image.

$$f_4 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p^2_{d,\vartheta^o}(i,j) \tag{5}$$

(V) *Homogeneity.* – Homogeneity is inversely proportional to contrast at constant energy whereas it is inversely proportional to energy (Park et al., 2002; Xian, 2010).

$$f_5 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{p_{d,\vartheta^o}(i,j)}{1+(i-j)^2} \tag{6}$$

Artificial neural network (ANN)

An ANN model is composed of an input layer, one or more hidden layers, and an output layer. Generally, the number of neurons in the input layer is equated to the input number in the problem while the number of neurons in the output layer is equated to the desired output number. The number of hidden layers and the neuron number in the hidden layers are determined by trials. The cells in the input layer are sent to the next layer without making any changes on the input. In the hidden layer, inputs and related weights are multiplied and the results are transmitted to transfer function (Rocha, 2007).

Neurons in each layer and weights that connect these to one another were used (Figure 3). In Figure 3, circles represent neurons while lines that bind neurons to one another represent weights. One of the most important things in an ANN is the bindings that provide data transmitted between neurons. A binding that transmits data from a neuron to another one has also a weight value. Thus, *G(x)* is a summation function and calculates the exact input

Table I. ANN training and testing parameters.

Number of layers	3
Number of neurons in layers	Input: 5
	Hidden: 10-50
	Output: 1
Initial weights and biases	Nguyen-Widrow Method
Activation functions	Tangent-sigmoid
Training parameters	
Learning rule	Backpropagation
Sum-squared error	0.0001

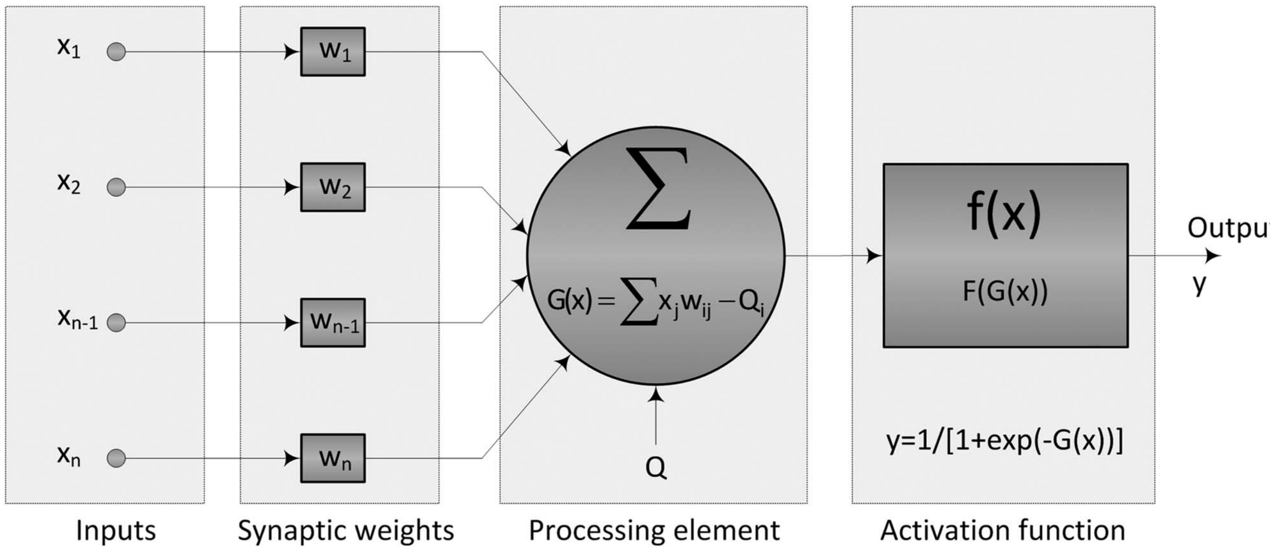

Figure 4. The model of an artificial neuron in the proposed ANN method.

that reaches a neuron (Figure 4). The input, by multiplying with variables and weight coefficients, builds up input for *G(x)* summation function (Figure 4). Mathematical statement of an artificial neuron can be written as:

$$y_i = F(G(x)) = F\left(\sum_{i=1}^{n} w_{ij}x_j - Q_i\right); x_i = (x_1, x_2, \ldots\ldots x_n) \quad (7)$$

In equation (7), $x = \{x_1, x_2, x_3, \ldots x^n\}$ is an input variable to be processed. On the other hand, $w = \{w^{00}, w^{01}, \ldots w^{ij}\}$ is the weight and it shows the importance of data incoming to a neuron and the impact on the neuron (Li & Flenley, 1999). The values of the weights can change at the process of training: Q_i represents the threshold value; *F(.)* is the activation function; *G(.)*, that comes to *F(.)*, is the function that produces the output by processing the input. There are different activation functions such as sigmoid, tangent sigmoid, sine and radial basis. In an ANN, all of the neurons may have the same or different activation functions. In order for many ANN models to make the calculations easier, activation functions are required of which derivates can be taken. Which activation function to be used is decided as a consequence of the user trials.

Results

GLCM texture features were obtained from the images of pollen in honey samples and the same procedures were applied to the pollen of selected plant species. The pollen data of plant species were compared with the pollen in honey (Figure 5). As a result of this study, identification and classification of the pollen type by using texture features achieved significant success.

The performance of an ANN depends on the number of neurons on the hidden layer and the activation function to be used. Hence, the appropriateness of the parameters has to be determined through a series of trials (Table I). As a consequence of this, activation functions such as sigmoid, tangent-sigmoid, sine and radial basis have been used for training and testing of the network. Numbers of neurons on the hidden layer have been found by being tested between ten and 50 after being increased one by one. The most appropriate activation function and neuron number have been settled according to the training and testing performance of the network. In the identification of the pollen grains type, the most appropriate activation function has been found as tangent-sigmoid.

We used ten different pollen types and five different image positions of these types for each plant species and honey. The aim of the study is to obtain an accurate identification of honey pollen by the reference to pollen from known plant species.

Table II. The performances of ANN for different direction and offsets.

Angle/offset	$d = 1$	$d = 2$	$d = 3$	$d = 4$
0°	78.00%	82.00%	76.00%	76.00%
45°	76.00%	80.00%	82.00%	76.00%
90°	68.00%	72.00%	72.00%	76.00%
135°	86.00%	78.00%	82.00%	82.00%

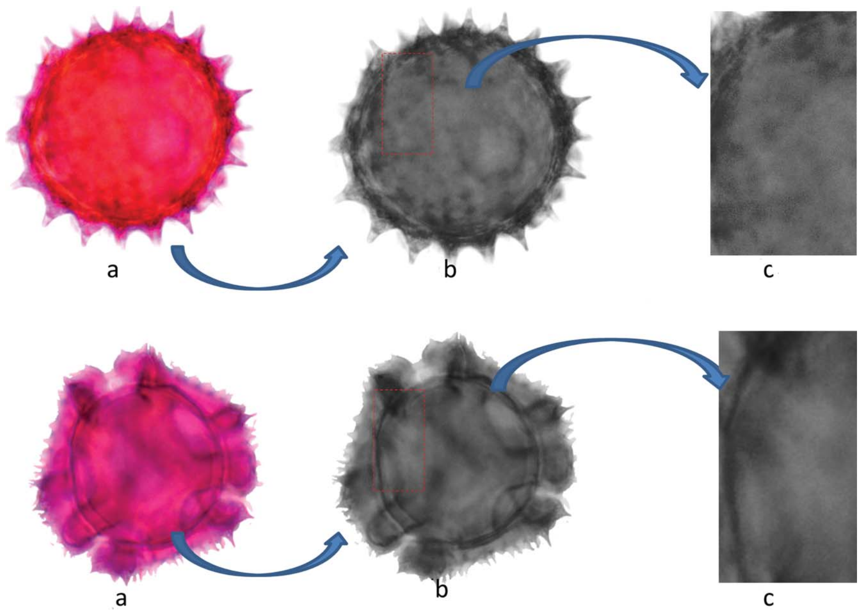

Figure 5. Demonstration of the mechanism: pollen images (**A**) are transformed into grey level images (**B**) of which the textures are identified in detailed enlargements (**C**).

The results of the tests using different angles and distances showed that the best results were obtained with texture features extracted at 135° from GLCM (Table II).

The success of 88.00% was obtained with the use of all features extracted at different angles and offsets. SVM, Naive Bayes, PART and Jrip machine learning methods applied to the data set were created for all different angles and offsets. Pollen identification successes by different machine learning methods vary with each other. The best result was obtained by ANN (Table III).

Discussion

Plants have many different pollen types and shapes, but also identification of pollen features is required in palynological studies, which have been performed in recent years. Therefore, rapid and accurate identification of pollen of the species are an important criterion for identification of the botanical origin of honey samples and consequently reduce costs associated with typing honey. In addition, it would also make the process more attractive for honey testing by agencies involved in detecting fraudulent labelling.

Table III. The success rates of different machine learning methods in pollen identification.

Method	Accuracy (%)
Naive-Bayes	72.00
SVM	72.00
PART	76.00
OneR	76.00
Jrip	76.00
ANN	88.00

Physico-chemical properties of honey samples vary from region to region and can be used to assess the quality. However, they do not provide any information about the origin of the honey and where it was produced. In multi-floral honey samples, we currently have to use palynological characteristics to identify the origin and a rapid and accurate automated technique may solve the problem for beekeepers and food monitoring agencies in having correct labels on honey sold commercially. While SEM photograph systems can achieve this, they are very expensive and require an expert system, identification of pollen with light microscopy and computer systems is much cheaper and more applicable.

In this study, ANN and GLCM methods were used for the identification of pollen grains from ten plant species and honey. Texture features are characters that specify the surface properties of the pollen. The surface textures of objects provide important advantages in pattern recognition or pattern matching. Texture features were obtained from GLCM. GLCM has been calculated in four angles and four offsets for pollen images. At the final stage, each angle and distance pair includes five features; ANN trained using training data set obtained from plants' pollen. The validity of the ANN model was tested with the data set that was created from honey pollen grains. ANN correctly identified the type of pollen for 44 of the 50 samples obtained from honey. The recognition rate of the model was 88%. Some special characters, such as angles and distances, which had the best classification results and highest percentages of accuracy, were additionally selected as useful features. These findings suggest that texture parameters can be very useful in identifying pollen types. As a result, the presented methodology effectively identified the pollen types subjected to in our study. In future studies, a database of honey plants could be generated, preferably for different or all countries. Improvement of these computer systems could be useful for the detection of the origin of all honey samples.

Conclusions

In this study, we aimed to design an effective, accurate, rapid and non-destructive computer vision method for pollen classification in honey, based on the textures of pollen images. As a result, the presented methodology effectively identified the pollen types subjected to in our study. In future studies, a database of honey plants could be generated, preferably for different or all countries. Improvement of these computer systems could be useful for the detection of the origin of all honey samples.

References

Bambara, S. B. (1991). Using pollen to identify honey. *American Bee Journal*, *131*, 242–243.

Benyon, F. H. L., Jones, A. S., Tovey, E. R. & Stone, G. (2000). Differentiation of allergenic fungal spores by image analysis, with application to aerobiological counts. *Aerobiologia*, *15*, 211–223.

Bolchi Serini, G. & Salvi, G. (1990). I mieli Lombardi: Rizultati di analisi palinologiche. *Ape Nostra Amica*, *12*, 6–9.

Davis, P. H. (1970). *Flora of Turkey and the East Aegean Islands*. Edinburgh: Edinburgh University Press.

Erdtman, G. (1969). *Handbook of palynology*. New York: Hafner.

Ferreira, I. C. F. R., Aires, E., Barreira, J. C. M. & Estevinho, L. M. (2009). Antioxidant activity of Portuguese honey samples: Different contributions of the entire honey and phenolic extract. *Food Chemistry*, *114*, 1438–1443.

France, A. W. G., Duller, G. A. T. & Duller, H. F. L. (2000). A new approach to automated pollen analysis. *Quaternary Science Reviews*, *19*, 537–546.

Guo, B., Damper, R. I. & Gunn, S. R. (2008). A fast separability-based feature-selection method for high-dimensional remotely sensed image classification. *Pattern Recognition*, *41*, 1653–1662.

Haralick, R. M., Shanmugam, K. & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics SMC*, *3*, 610–621.

Hermosin, I., Chicon, R. M. & Dolores Cabezudo, M. (2003). Free amino acid composition and botanical origin of honey. *Food Chemistry*, *83*, 263–268.

Huang, K. (2007). Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. *Computers and Electronics in Agriculture*, *57*, 3–11.

Jones, A. S. (2000). Image analysis applied for aerobiology. *Proceedings of the 2nd European Symposium on Aerobiology, Vienna, Austria, 5–9 September 2000*, 2.

Kaya, Y., Yesilova, A. & Almalı, M. N. (2010). An application of expection and maximization, multiple imputation and neural network for missing value. *World Applied Sciences Journal*, *9*, 561–566.

Langford, M., Taylor, G. E. & Flenley, J. R. (1990). Computerised identification of pollen grains by texture. *Review of Palaeobotany and Palynology*, *64*, 197–203.

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

Louveaux, J. (1985). *Les abeilles et leur élevage*. Echaufour: OPIDA.

Ondimu, S. N. & Murase, H. (2008). Effect of probability-distance based Markovian texture extraction on discrimination in biological imaging. *Computers and Electronics in Agriculture*, *63*, 2–12.

Park, B., Lawrence, K. C., Windham, W. R., Chen, Y.-R. & Chao, K. (2002). Discriminant analysis of dual-wavelength spectral images for classification poultry carcasses. *Computers and Electronics in Agriculture*, *33*, 219–231.

Piana, L., Oddo, L. P., Bentabol, A., Bruneau, E., Bogdanov, S. & Declerck, C. G. (2004). Sensory analysis applied to honey: State of the art. *Apidologie*, *35*, 26–37.

Rocha, M. (2007). Evolution of neural network for classification and regression. *Neurocomputing*, *70*, 2809–2816.

Ronneberger, O. (2000). Automated pollen recognition using gray scale invariants on 3D volume image data. *Proceedings of the 2nd European Symposium on Aerobiology, Vienna, Austria, 5–9 September 2000*, 3.

Sato, T. & Miyata, G. (2000). The nutraceutical benefit, part II: Honey. *Nutrition*, *16*, 468–469.

Szczęsna, T. & Rybak-Chmielewska, H. (1993). Analiza pyłkowa miodów (przegląd literatury). *Pszczelarstwo*, *5*, 10–11 (in Polish).

Vezey, E. L. & Skvarla J. J. (1990). Computerised feature analysis of exine sculpture patterns. *Review of Palaeobotany and Palynology*, *15*, 187–196.

von der Ohe, W. (1994) Unifloral honeys: Chemical conversion and pollen reduction. *Grana*, *33*, 292–294.

Wodehouse, R. P. (1935). *Pollen grains*. New York: Hafner.

Xian, G. (2010). An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM. *Expert Systems with Applications*, *37*, 6737–6741.