

Discrimination of bio-crystallogram images using neural networks

Sevcan Unluturk · Mehmet S. Unluturk ·
Fikret Pazir · Alper Kuscü

Received: 29 February 2012 / Accepted: 25 January 2013 / Published online: 9 February 2013
© Springer-Verlag London 2013

Abstract This study utilized a unique neural network model for texture image analysis to differentiate the crystallograms from pairs of fresh red pepper fruits from conventional and organic farms. The differences in visually analyzed samples are defined as the distribution of crystals on the circular glass underlay, the thin or thick structure of crystal needles, the angles between branches and side needles, etc. However, the visual description and definition of bio-crystallogram images has major disadvantages. A novel methodology called an image neural network (INN) has been developed to overcome these shortcomings. The $1,488 \times 2,240$ pixel bio-crystallogram images were acquired in a lab and cropped to $425 \times 1,025$ pixel images. These depicted either a conventional sweet red pepper or an organic sweet red pepper. A set of 19 images was utilized to train the image neural network. A new set of 4 images was then prepared to test the INN performance. Overall, the INN achieved an average recognition performance of 100 %. This high level of recognition suggests that the INN is a promising method for the discrimination

of bio-crystallogram images. In addition, Hinton diagrams were utilized to display the optimality of the INN weights.

Keywords Back propagation learning algorithm · Bayes optimal decision rule · Gram–Charlier series · Hinton diagrams · Neural network · Probability density function · Bio-crystallogram images · VLSI

1 Introduction

Excluding analytical chemistry approaches for the evaluation of food quality, there are numerous alternative methods (holistic methods) which, rather than focusing on the analysis of single substances, regard all foodstuffs as part of a whole. This approach is based on the proposition that life itself is more than the sum of its single constituent parts. The organization and structure of food are important factors. The conservation of this organization and structure is a sign of high food quality, and structural decomposition is synonymous with plant death. The purpose of complementary analysis methods is to help characterize the vital activity [1]. The bio-crystallization method is one of the most important complementary analysis methods.

Pfeiffer (1931) originally introduced the term bio-crystallization, which is also called “sensitive crystallization” and “copper chloride crystallization.” Engqvist (1970) [2, 3] later initiated the bio-crystallization method. This method was developed from the viewpoint that living organisms not only exist as substances, but have directing and organizing “structuring forces.” These structuring forces direct the form and function of the organism [4]. This method is based on the following crystallographic phenomenon; the addition of specific ionic substances, or in general, all organic substances to an aqueous solution of

S. Unluturk
Food Engineering Department, Izmir Institute of Technology,
Izmir, Turkey

M. S. Unluturk (✉)
Department of Software Engineering, Izmir University
of Economics, Sakarya Cad. No. 156, 35330 Balcova,
Izmir, Turkey
e-mail: suleyman.unluturk@ieu.edu.tr

F. Pazir
Food Engineering Department, Ege University, Izmir, Turkey

A. Kuscü
Faculty of Agriculture, Suleyman Demirel University,
Isparta, Turkey

dehydrate CuCl_2 causes crystallization, during which bio-crystallogram with reproducible dendritic structures are formed. Crystallograms that are produced on the basis of pure CuCl_2 exhibit a merely peripheral distribution of crystals on the circular glass underlay, with a diameter of 90 mm [5]. In contrast, bio-crystallograms produced on the basis of biological substances, such as plant extracts (fresh sweet red pepper samples in this study), exhibit crystal structures covering the entire glass underlay (see Fig. 1).

The bio-crystallograms produced from agricultural products, such as vegetables, grains, fruits, and milk samples are based on three components: (a) an aqueous solution or extract of the sample in question; (b) an aqueous solution of dehydrate copper chloride; and (c) purified water. Any type of additive will change the copper chloride crystallization. The process is influenced by the qualitative and quantitative variations in the macromolecules of the biological extracts, thus allowing food quality assessment. When used to study human blood, the results are correlated with groups of pathologies and the concentration of blood protein and are predictive of worsening illnesses. Early

diagnosis seems possible in cancer research or in occupational medicine (risk indicators) [6].

The phenomena of bio-crystallograms are based on ramification patterns, which may be divided into three major stages, extending from the center in all directions to the periphery of the image (see Fig. 2). In the first or 1-zone bio-crystallogram, when increasing concentrations of agricultural extracts are applied relative to a given fixed concentration of CuCl_2 , transparent needles with enormous starlike formations extend in all directions to the periphery. The second, 2-zone structure goes through the crystallization center when it is divided by a vertical and horizontal axis. The needles are pointed, predominantly transparent, and of relatively equal length in the middle zone. These morphological features may be described by means of plant morphological terms, such as stems, branches, and needles. The last stage of a bio-crystallogram is an optimal degree, divided into a 3-zone structure with a center zone, middle zone, and marginal zone. In the third stage, a bio-crystallogram exhibits various macro- and microscopic morphological features that reflect the quality of the sample in question [6].

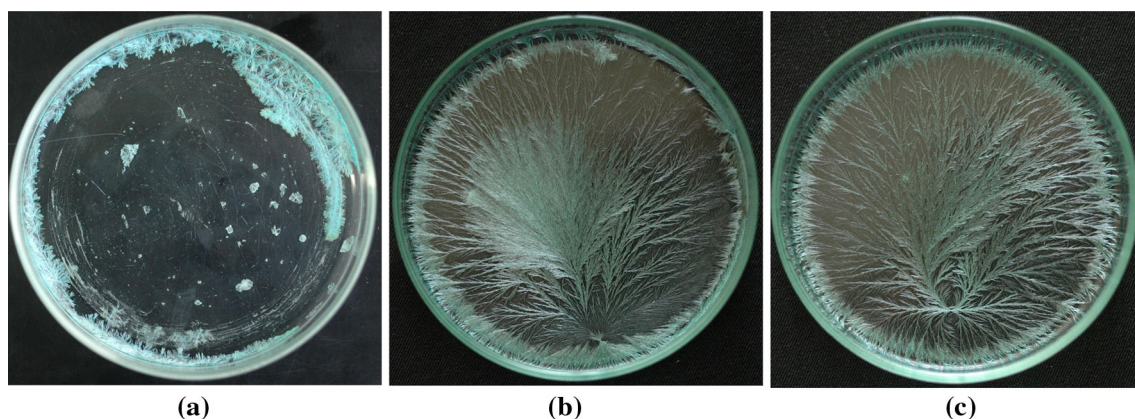


Fig. 1 **a** Crystallogram obtained from basis of aqueous dehydrate CuCl_2 (blank), **b** conventionally growing fresh sweet red pepper bio-crystallogram, **c** organically grown fresh sweet red pepper bio-crystallogram [7]

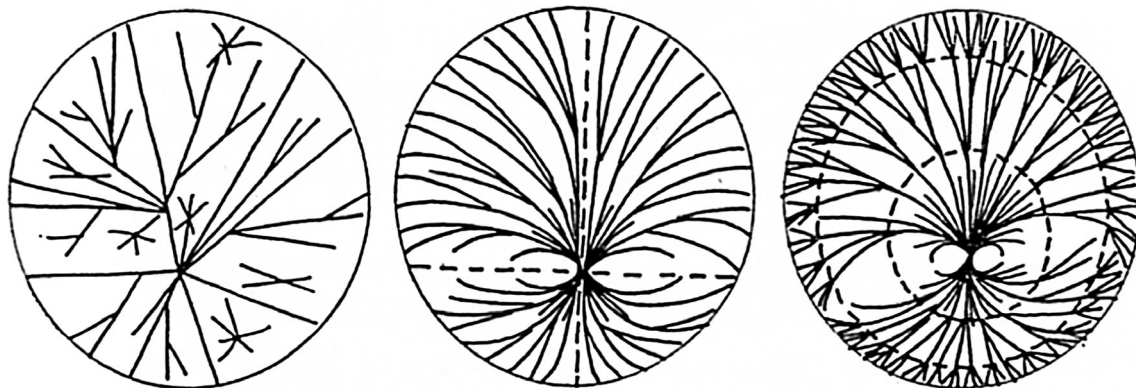


Fig. 2 A Bio-crystallization method based on a phenomenon of dendritic pattern formation during crystallization from an aqueous solution containing plant extracts and CuCl_2 [2]

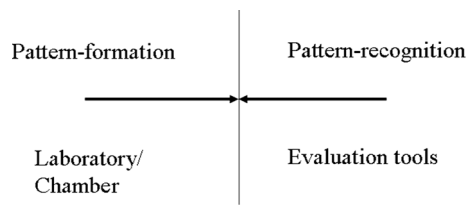


Fig. 3 The detail of bio-crystallization method [8]

The bio-crystallization method comprises two main parts (see Fig. 3). The first is pattern formation, which starts in a laboratory and continues until the crystallogram picture is completely generated in the chamber. The second part is pattern recognition. This is where the evaluation tools will be responsible for perceiving and differentiating between images [3].

There are two tools currently used to evaluate an image, visual evaluation, and computerized image analysis. In visual evaluation, the images in question are evaluated based on the judgment of a trained human using discrete reference scales arranged in connection with picture phenomena. In contrast, computerized image analysis interprets the image by using the fundamental knowledge of texture analysis. Such techniques have been explored and applied with the bio-crystallization method [6].

Computerized image analysis techniques may be best able to meet the demand for such methods. Ideally, an image analysis procedure should reflect all of the characteristics of a bio-crystallogram as a three-dimensional, colored ramification structure, coordinated with zones relative to the center. However, due to the present limitations set by computational capacity and speed, simpler approaches are preferable. In a limited number of previous studies, encouraging results have been reported [5].

Computerized image analysis tools increase the objectivity of the method and allow the analysis of large numbers of crystallization images [8]. This paper presents a unique neural network model, called an image neural network, as an effective analysis tool.

2 Sweet red pepper extract and copper chloride concentrations

Conventional and organic sweet red pepper fruit was obtained from Ege University Agricultural Faculty Research, Application and Production Farm in Menemen (2005–2006; Izmir, Turkey). These fruit were transferred directly to the laboratory, where they were cleaned, finally chopped, and passed through a kitchen-type blender (Braun MR 404). The homogenate was first passed through a cloth to remove debris particles and then filtered. The fruit juice was diluted to 1 % with bi-distilled water. Copper chloride

dehydrate solution was also prepared at a 16 % concentration with bi-distilled water.

The optimal mixing ratio for the sample extract and copper chloride influences the crystallization pattern. Therefore, the optimum sample and CuCl_2 dehydrate concentrations were determined to be 1 and 16 %, respectively.

For each bio-crystallogram, a 1-ml sample (1 % concentration) was mixed with 3 ml CuCl_2 dehydrate (16 % concentration) per plate at chamber conditions of 25 °C and 55 % relative humidity [9, 10]. The copper chloride was crystallized after 16–18 h. A total of 23 bio-crystallograms were produced from organic and conventional fresh sweet red pepper fruit in 2005 and 2006. Single centered bio-crystallograms only were evaluated, and the others (multi-centered) were discarded.

3 Bayes optimal decision rule

All the algorithms for preprocessing of full images and image segmentation were written in MATLAB 6.5 (The Math Works, Inc., MA, USA). MATLAB code for preprocessing and image segmentation of a single full image is given below:

```
Z = imread('DSC_1851.JPG'); level = graythresh(Z);
BW = im2bw(Z, level);
```

where imread MATLAB function reads a color image from the file called 'DSC_1851.JPG'. The graythresh MATLAB function computes a threshold, level, that can be used to convert a color image represented by Z to a binary image using im2bw function.

Figure 4 shows a $1,488 \times 2,240$ pixel conventional single centered bio-crystallogram for a red pepper. This image was then transformed into a black and white image and cropped (see Fig. 5). Let us represent this new black and white image as a row vector, obtained from the image matrix by combining its rows:

$$\vec{X} = [x(1), x(2), \dots, x(m)]$$

where m is 435,625. Each element of the vector corresponds to a black and white color value that is either 0 or 1. Then, there are two classes from which our input image might be drawn.

These classes are defined as:

$$H_0 = \text{Organic}, \\ H_1 = \text{Conventional}.$$

The prior probability of an unknown image being drawn from class k is h_k ($k = 0, 1$). The cost of making a wrong decision for class k is v_k . Note that the prior probability h_k and the cost probability v_k are taken as being equal and



Fig. 4 Conventional red pepper image



Fig. 5 Cropped black and white red pepper image (425 pixel × 1,025 pixel)

hence can be ignored. The problem is to find a neural network model for determining the class from which an unknown image is taken.

If we know the probability density functions $f_k(\vec{X})$ for all classes, the Bayes optimal decision rule [11] can be used to classify \vec{X} into class k if

$$h_k \vartheta_k f_k(\vec{X}) > h_m \vartheta_m f_m(\vec{X}) \tag{1}$$

where $k \neq m$.

A major problem with Eq. (1) is that the probability density functions of the classes are unknown; however, these can be estimated using Gram–Charlier series expansion. In this case, the training set for the neural network consists of Gram–Charlier coefficients. Our objective in this study was to use these coefficients for classification. If the neural network is trained with known Gram–Charlier coefficients, then it is only necessary to feed its Gram–Charlier coefficients to determine the class of an unknown image. The next section describes the Gram–Charlier series expansion.

4 Gram–Charlier series

The Gram–Charlier series expansion of the probability density function of a random variable with mean μ and variance σ^2 can be represented as follows:

$$\rho(x) = \frac{1}{\sigma} \sum_{i=0}^{\infty} c_i \phi^{(i)}\left(\frac{x-\mu}{\sigma}\right), \tag{2}$$

where $\phi(x)$ is a Gaussian probability density function and $\phi^{(i)}(x)$ represents the i -th derivative of $\phi(x)$. For normalized data where $(\mu = 0, \sigma^2 = 1, c_0 = 1)$, the above equation can be simplified to

$$\rho(x) = (\phi(x) + c_3\phi^{(3)}(x) + c_4\phi^{(4)}(x) + c_5\phi^{(5)}(x) + c_6\phi^{(6)}(x) + \dots), \tag{3}$$

where c_i coefficients are related to the central moments of $\phi(x)$. In this sense, derivatives of the Gaussian function in Eq. (3) provide us with the class type information for the sweet red pepper. Furthermore, $c_i\phi^{(i)}$ are orthogonal functions that present unique information about the organic and conventional pepper class distribution. This leads us to conclude that an INN based on the decomposition of the probability density function by the Gram–Charlier series is well suited for organic/conventional pepper class discrimination. Let us define $\beta(x)$ as

$$\beta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \tag{4}$$

Then, the probability density function becomes

$$\rho(x) = \beta(x) \left\{ 1 + \frac{1}{6} \Lambda_3 A_3 + \frac{1}{24} (\Lambda_4 - 3) A_4 + \dots \right\}, \tag{5}$$

where

$$A_i(x) = x^i - \frac{i^{[2]}}{2 \cdot 1!} x^{i-2} + \frac{i^{[4]}}{2^2 \cdot 2!} x^{i-4} - \frac{i^{[6]}}{2^3 \cdot 3!} x^{i-6} + \dots, \tag{6}$$

$$i^{[k]} = \binom{i}{k} = \frac{i!}{(i-k)!}$$

$$m_k = \frac{1}{n} \sum_{i=1}^n [x(i)]^k$$

$$\Lambda_3 = -\frac{m_3 - 3m_2m_1 + 2m_1^3}{3!},$$

$$\Lambda_4 = \frac{m_4 - 4m_3m_1 + 6m_2m_1^2 - 3m_1^4}{4!},$$

$$\Lambda_5 = -\frac{m_5 - 5m_4m_1 + 10m_3m_1^2 - 10m_2m_1^3 + 4m_1^5}{5!},$$

$$\Lambda_6 = \frac{m_6 - 6m_5m_1 + 15m_4m_1^2 - 20m_3m_1^3 + 15m_2m_1^4 - 5m_1^6}{6!}$$

Equation (6) is the so-called Gram–Charlier series [12] and the polynomial $A_i(x)$ is called the Tchebycheff–Hermite polynomial.

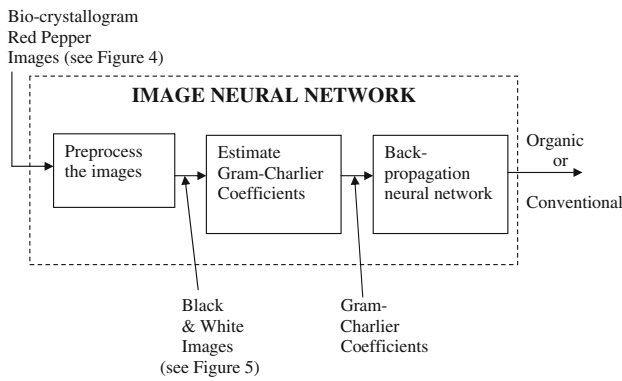


Fig. 6 Image neural network (INN)

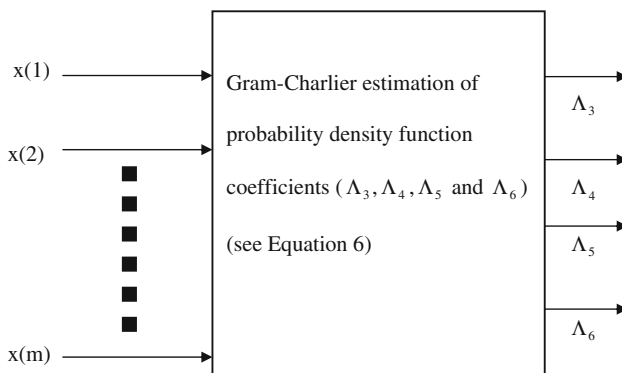


Fig. 7 Cropped black and white 425 × 1,025 pixel image is converted to a 1 × 435,625 pixel row vector \vec{X} (where m is 435,625). This row vector is normalized before its Gram–Charlier coefficients are calculated ($x(i), i = 1..m$ is normalized (the mean value is removed and the standard deviation is set to be 1)

5 Image neural network architecture

This novel neural network has three processing stages (see Fig. 6). The first stage preprocesses bio-crystallogram red pepper single centered images, and the second stage converts these black and white images into row vectors and estimates Gram–Charlier coefficients [11] corresponding to these row vectors (see Fig. 7). In the third processing block, these parameters are applied to a back propagation neural network in order to classify the images (see Fig. 8). This neural network is a fully connected feed forward neural network. When the Gram–Charlier coefficients are presented as an input to the neural network, each output of the hidden neurons is a weighted sum of the input nodes and the bias node passes through a hyperbolic tangent function. The hidden neuron’s output of this neural network, is given as

$$y_j = \phi \left(\sum_{i=1}^M x(i)w_{ji}^h + \theta^j \right) \tag{7}$$

where L is the number of hidden neurons and is taken as 10, M is number of inputs and is taken as 4, and the activation function for the hidden and output layer, $\phi()$, is a tangent hyperbolic function defined as

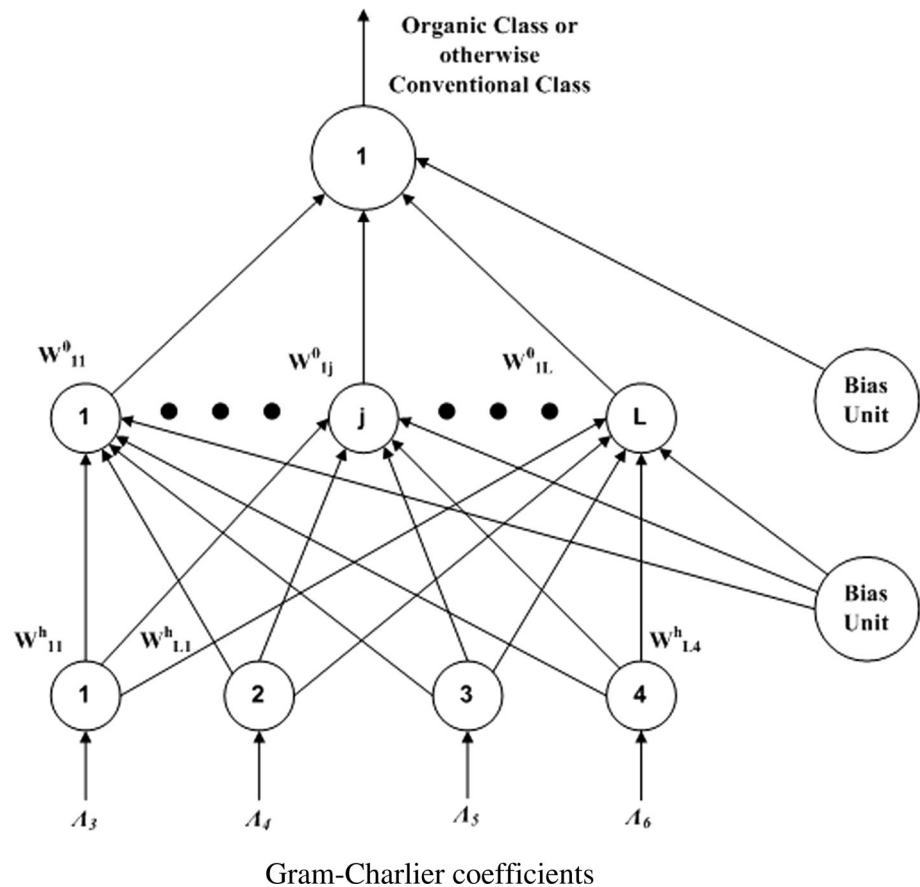
$$\phi(y) = \frac{e^y - e^{-y}}{e^y + e^{-y}} \tag{8}$$

The term is the input vector that consists of the Gram–Charlier coefficients and is the weight from the hidden neuron to the input neuron. Then, the output of the neural network is the weighted sum of all the hidden neurons and the bias node. This actual output of the neural network is then compared to the desired output for every set of input values. Furthermore, the neural network error is defined as the difference in these outputs. Then, the weights of the neural network are updated using the gradient of this output error [13, 14].

The training matrix was prepared using the Gram–Charlier coefficients. Each column in the training matrix represented one set of these Gram–Charlier coefficients with a length of 4. Note that there could be any number of columns in the training set. The greater the number, the better the discrimination achieved by the INN for the target bio-crystallogram red pepper single centered images.

The results obtained when testing the INN show that bio-crystallograms of organic and conventional sweet red pepper targets can be detected with 100 % accuracy. Out of the 23 images, 19 were used for training where 10 of them belong to the Organic class and 9 of them belong to the Conventional class. A total of 4 images were used for testing. Two of them belong to the Organic class, and two of them belong to the Conventional class. The number of hidden neurons for the back propagation neural network was chosen to be 10, based on experience. In testing, if the output of the INN ≥ 0 , then the input Gram–Charlier coefficients belonged to class $H_o = \text{Organic}$. If the output of the INN < 0 , then the input Gram–Charlier coefficients belonged to class $H_1 = \text{Conventional}$ (see Fig. 8). The success of the INN may be obscured if there is no understanding of the optimality of the neural network. In this study, Hinton diagrams were used to explain the optimal characteristics of the neural network. Hinton diagrams are named after Geoffrey Hinton, who used this method to display neural network weights [15]. In the Hinton diagrams, the size of each square is proportional to the strength of each weight. Green and red represent positive and negative magnitude weights, respectively. Figure 9a depicts the Hinton diagram of the INN. The first row shows the weights from the input nodes to the first hidden neuron, the second row depicts the weights from the input nodes to the second hidden neuron, and so on. Note that the number of rows in Hinton diagrams is equal to the number of

Fig. 8 Back propagation neural network. If the output of the neural network ≥ 0 , then input type of bio-crystallogram sample image belongs to organic class, otherwise it belongs to conventional class



hidden neurons. The first column shows the weights between each hidden neuron and the first input neuron. Figure 9b depicts the weights from the hidden neurons to the output nodes. Since there is one output node, the number of rows in this Hinton diagram is 1. Again, the first column depicts the weights between the output nodes and the first hidden neuron.

When we examine the Fig. 9a, b, the hidden neurons that have smaller weights have also smaller weight to the output neuron; on the other hand, the hidden neurons that have bigger weights have also bigger weight to the output neuron. Hence, all the hidden nodes contribute to the decision, and we do not have a small weight in the output layer where it cancels a bigger weight in the input layer.

Performance of this INN is evaluated using the data in matrix given below. The following table shows the confusion matrix for Organic-Conventional INN classifier:

When the Table 1 is examined, one can see that the INN is not confusing the two classes, namely Organic and Conventional.

This novel neural network has the following properties: (1) the response is real-time once it is trained; (2) all hidden neurons have the same structure, which allows convenient hardware implementation using VLSI design

techniques [16, 17]. Based on experimental observations, the INN is effective in the discrimination of the crystallograms of conventionally and organically grown fresh red pepper fruits.

6 Conclusion

In this study, we developed a neural network designed to classify the bio-crystallograms of conventional and organic red peppers. The images taken from a lab bear information related to the class type of pepper. However, this information is not readily quantifiable and lacks uniquely recognizable features. Therefore, because it is trainable, a neural network becomes appealing for classifying these images. The optimal values for the neural network weights were estimated using the back propagation algorithm. Experimental measurements of the pepper were utilized to train and test the image neural network. This network showed a remarkable 100 % success in classification performance. Parallel classification performance was also achieved when training the neural network. These encouraging results suggest that neural networks are potentially useful for detecting conventional/organic red peppers and in quality

Fig. 9 **a** Hinton diagram displays the estimated hidden layer weights of the INN. **b** Hinton diagram displays the estimated output weights of the INN

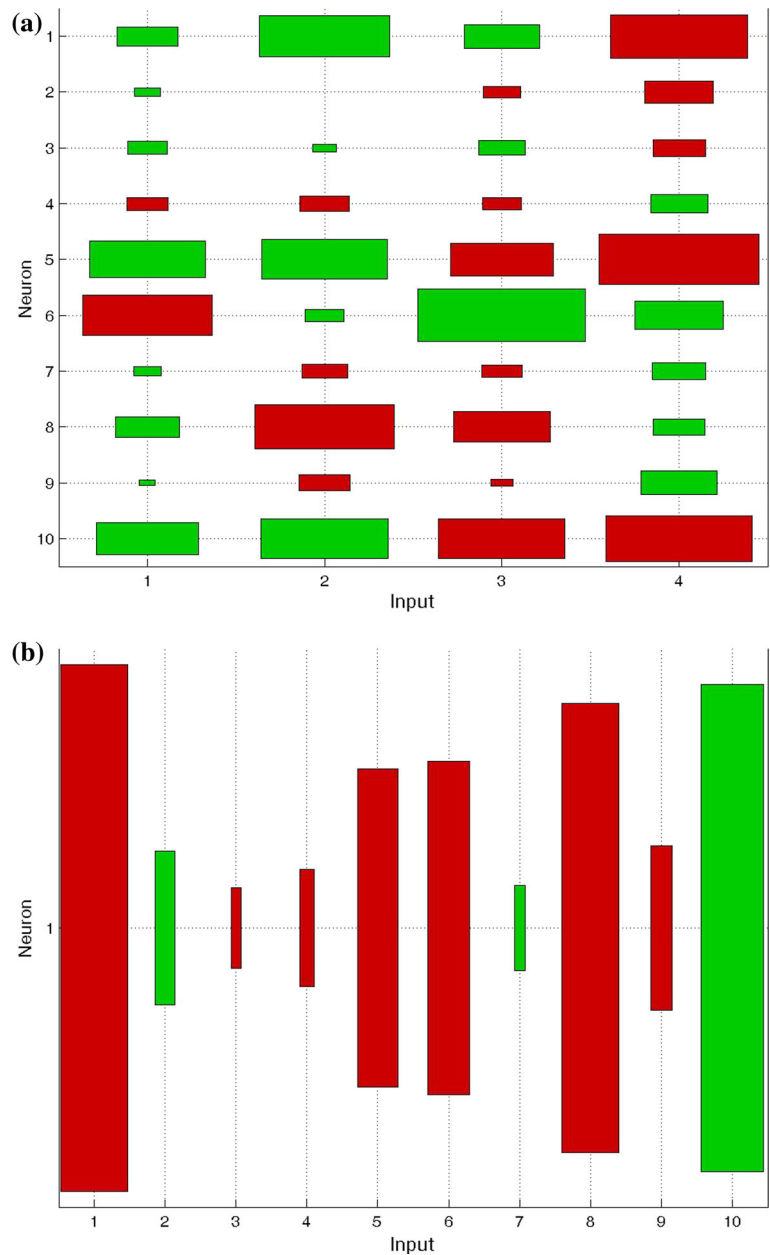


Table 1 Confusion matrix for testing results of two class classifier, INN

Actual	Predicted	
	Negative	Positive
Negative	2	0
Positive	0	2

control. Furthermore, the image neural network renders practical advantages such as real-time processing, adaptability, and training capability. Another potentially important role for similar neural network designs is medical ultrasonic imaging for tissue characterization and diagnosis.

7 Future work

The Bio-crystallization method can be a low cost and innovative method to discriminate samples of different origin (e.g., organic vs. conventional). One possible application for this Neural Network image processing for the classification of bio-crystallograms is that it could be adapted to develop a fingerprinting method for organic food authentication. This would be promising as there is no scientific literature available for this topic. Lastly, although the proposed method shows considerable potential for the classification of red pepper samples, it would be valuable to conduct further research using more varieties of red

peppers that have been sourced from different locations to investigate the impact this may have on the classification results.

References

- Harms W (2004) Quality of organic food. *Genet Eng Newslett (Special Issue)* 16:1–11
- Pfeiffer E (1931) Studium von Formkräften an Kristallizationen. Naturwissenschaftliche Sektion am Goetheanum, Dornach, Switzerland
- Engqvist M (1970) Gestaltkräfte des Lebendigen. Klostermann, Frankfurt amMain, Germany
- Bloksma J, Northolt M, Huber M (2001) Parameters for apple quality part-1. Louis Bolk Institute, p 115 (Archived at <http://orgprint.org/4266>)
- Andersen JO, Henriksen CB, Laursen J, Nielsen AA (1999) Computerised image analysis of biocrystallograms originating from agricultural products. *Comput Electron Agric* 22(1):51–69
- Meelursam A (2006) Statistical evaluation of texture analysis from the biocrystallization method: effect of image parameters to differentiate samples from different farming systems, submitted in fulfillment for the degree of doctor of agricultural science. University of Kassel, Department of Organic Food Quality and Food Culture, Witzenhausen, December 2006, p 220
- Kuşçu A (2008) Organik ve Konvansiyonel Kırmızıbiber ve Ürünlerinin Ayırt Edilebilme Yöntemleri ve Kalite Özelliklerinin İncelenmesi. Ege Üniversitesi Fen Bilimleri Enstitüsü, Ph.D. Thesis, 420 p, Izmir
- Doesburg P, Huber M (2007) Biocrystallisation and Steigbild results at the Louis Bolk Instituut. *Elemente Der Naturwissenschaft* 87:118–123
- Balzer-Graf UR, Köpke U, Geier U (2001) Research on quality in organic agriculture by picture forming methods, understanding the quality of organic horticultural products (Short Course on). 15–26 May, Izmir, Turkey
- Szulc M, Cordeiro F, Maquet A, Anklam E (2005) Application of a multivariate design approach for maximisation of the observed differences between organically and conventionally grown wheat grains in biocrystallisation method. In: Proceedings of the 1st scientific FQH conference, November 28th & 29th, Fibl, Frick, p 78
- Timothy Masters (1993) Practical neural networks. Academic Press Inc., Waltham, Massachusetts
- Kendall MG, Stuart A (1963) The advanced theory of statistics. Charles Griffin & Company Limited, London
- Cichocki A, Unbehauen R (1992) Neural networks for optimization and signal processing. Wiley, Hoboken
- Freeman JA, Skapura DM (1991) Neural networks algorithms, applications, and programming techniques. Addison-Wesley Publishing Company, Boston
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by backpropagation error. *Nature* 323:533–536
- Hollis PW, Paulos JJ (1988) Artificial neural networks using MOS analog multipliers. In: International conference on neural networks
- Hollis PW, Paulos JJ (1994) A neural network learning algorithm tailored for VLSI implementation. In: IEEE transactions on neural networks, vol. 5, No. 5