International Journal of Horticultural Science and Technology
Journal homepage: https://ijhst.ut.ac.ir

Medicinal Plants Authenticity Evaluation using an Intelligent Hyperspectral Imaging System Coupled with Biologically Inspired Unsupervised Algorithms: *Nepeta Crispa* Willd Case Study

Sajad Kiania1*, Mahdi Ayyari2*

1 Biosystems Engineering Department, Sari Agricultural Sciences and Natural Resources University, Sari, Iran
2 Horticultural Science Department, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

*Corresponding author's email: s.kiani@sanru.ac.ir, m.ayari@modares.ac.ir

Article history:

Received: 29 October 2024,
Received in revised form: 12 April 2025,
Accepted: 25 April 2025,

Article type:

Research paper

Keywords:

Artificial Neural Networks,
Clustering,
Geographical Origins,
Medicinal Plants

ABSTRACT

This study evaluated the efficiency of a handheld hyperspectral Imaging (HSI) system coupled with biologically inspired unsupervised algorithms as a screening technique for the authenticity evaluation of naturally-grown *Nepeta crispa* Willd (*N. crispa*) samples from on-farm cultivated samples. The volatile oils of 25 samples were isolated with hydrodistillation, and then the Gas Chromatography (GC) analyses were done to determine the total Volatile Organic Compounds (VOCs) as the reference data. On the other hand, the samples' reflectance spectra were captured using the HSI camera and pre-processed using the Savitzky-Golay (SG) algorithm. Principal Component Analysis (PCA) was then applied for the visual discrimination of the samples and the data reduction. Next, two unsupervised algorithms, crisp clustering by the self-organizing Map (SOM) and an automatic clustering based on the artificial bee colony (ABC), were applied to perform the real clustering of the samples. The SOM unified distance matrices explained the changes in spectral characteristics between the *N. crispa* samples and indicated the variation of the sample's VOCs following the GC results. The automatic clustering by the ABC algorithm illustrated its capability to cluster the two main sample groups according to the sample's spectra. The HSI system combined with the ABC algorithm will provide a novel nondestructive and rapid technique for evaluating the authenticity of naturally-grown *Nepeta crispa* Willd samples.

Introduction

Nepeta crispa Willd. (*N. crispa*) is one of the best well-known medicinal and aromatic plants in the *Nepeta* genus. This plant has aerial parts with a sweet odor which is traditionally used for making an infusion (herbal tea), floral water, and different beverages. Due to its many nutritional and medicinal properties such as stomach pain relief, febrifuge, sedative, relaxant, carminative, and restorative tonic for nervous and respiratory disorders, it has been used in phytotherapy (Sefidkon et al., 2006). The extract of *N. crispa* is also used in ointments to heal skin disorders of eczema type and as a diuretic with slight bacteriostatic activity (Mojab et al., 2009). Therefore, the consumption of *N. crispa* and its products has gradually increased over the last decades. Due to its striking market, it has been cultivated on farms in different geographical origins besides its natural production. Many studies have proved that the quality of medicinal plants is determined by their total VOCs (Rehman et al., 2020). Although the composition of essential oils of medicinal plants and the percentage of total VOCs, are mainly made by directing genetic processes, but are significantly influenced by environmental factors resulting from geographical origins and cultivation conditions (Nejadhabibvash et al., 2018). Cultivation

conditions such as differences in height, rainfall, soil type, and environmental stresses usually affect the quantity and quality of *N. crispa* plant active ingredients. The percentage of the VOCs of *N. crispa* samples that have been naturally grown and harvested in mountainous areas is much more than samples harvested from farms (Karami et al., 2020). In the *N. crispa* market, consumers know that natural products are superior in quality and seeking them. Both types of products are almost similar and cannot be easily categorized by human vision and smell power except by doing lab-based analyses such as gas chromatography (GC) or high-performance liquid chromatography (HPLC). Consequently, there is a possibility that low-quality products will be marketed as natural or high-quality products. Thus, it is important to pay more attention to inspecting and monitoring the distribution chain of *N. crispa*.

During the past three decades, different analytical tools and methods have been developed for the quality control of medicinal plants. These methods are based on GC or liquid chromatography (LC) (Petrakis et al., 2017). Although these conventional analytical methods are accurate and readily available, they are often expensive, time-consuming, and require professional knowledge of the operation. Also, these systems cannot be used outside of the laboratory, industrially, and in real-time, especially in the retail markets during the distribution of the products (Kiani et al., 2018 and 2023). HSI is a type of screening technique that represents a perfect combination of conventional spectroscopy and regular imaging. It can simultaneously provide both spectral and visual information about each pixel in the image of the sample (Qin et al., 2013). The spectral data have the potential to extract detailed internal and external quality information and the chemical composition of agro-products. Many research studies have proved the capability of the HSI system for evaluating plants and spices. Some of the newest documented estimating nitrogen content of lettuce leaves (Odabas et al., 2017); measuring water content level on leaves and salt stress tolerances of soybean leaves (Sytar et al., 2017); identification of papaya seeds in black pepper (Orrillo et al., 2019); characterization of fermented cocoa beans from different origins (Acierno et al., 2019); Nutmeg authenticity evaluation (Kiani et al., 2019); Non-destructive determination of volatile oil and moisture content and discrimination of geographical origins of *Zanthoxylum bungeanum* Maxim (Ke et al., 2020); prediction of the oleic acid content of rapeseed (Liu et al., 2021); rapid detection of the nutrient content of hydroponically grown lettuce cultivars (Eshkabilov et al., 2021); predicting micronutrients of wheat (Hu et al., 2021), prediction of essential oil content in spearmint (Van Haute et al., 2023); Geographical origin differentiation and quality determination of saffron (Kiani et al., 2023); and rice authenticity evaluation (Edris et al., 2024 and 2025). Given the proven benefits of the HSI system, it can be stated that HSI technology is effective and applicable to the medicinal and aromatic plant industries. In HSI applications, supervised algorithms are more common for identifying the sample's varieties and authenticity or quality evaluation. The supervised methods use several labeled samples to train the algorithms and make a descriptor map between the spectral fingerprints and their identified characteristics. In these methods, the developed models are created to predict unknown sample properties. Using the constructed models at different times deals with two big challenges because training the models with spectral information that indicates all the pertinent characteristics is difficult. First, the supervised models might be affected by the sensitivity of the HSI camera sensors or illumination intensity and they might not be reproducible in responses for a target sample at two times. Second, in real applications, the system might face unknown samples with unknown patterns that the models were not trained for. Hence, the supervised model requires to be renewed frequently. Furthermore, applying different postharvest processes including processes from transportation to storage could affect N. crispa quality and of course its spectra. To solve this problem, unsupervised approaches could act as a capable tool for evaluating the samples. The unsupervised algorithms perform the real clustering of the spectral patterns only based on the discovery of undefined input data at the moment (Jose-Garcıa and Gomez-Flores, 2016). These algorithms are known for less intricacy and are fast because all the spectral fingerprints (input data) are fed to the models at once and grouping is done in real-time. This study aimed to propose a reliable screening approach to discriminate natural *N. crispa* from the samples cultivated on the farms. In this regard, hyperspectral images of the *N. crispa* samples were captured, preprocessed, reduced, and processed using two bio-inspired unsupervised algorithms.

Materials and Methods

Plant materials and reference values measurement

Fifty grams of aerial parts of 25 *N. crispa* samples were collected in the full flowering stage from three origins where it grows naturally and from two farms (5 samples from each origin/farm). Naturally grown samples were procured from three mountainous areas (Gashani, Arzanfood, and Pisteski) from Hamedan province, Iran. This province is the natural habitat of the plant. Other samples were cultivated and harvested on two farms (planted in similar growing conditions except the climate) in Tehran and Hamedan provinces, Iran. The samples were

harvested at a similar date and then dried in the shade to reach down to 10% moisture content (dry-based, $MC_{d.b}$). The samples were kept at a constant temperature of 4 °C before the measurements. The volatile oils of all samples were isolated with hydrodistilation, GC analyses were done, and the total VOCs of the samples, as well as components of oils, were determined. A detailed description of the GC analyses of the samples was provided by Karami et al. (2021). Table 1 shows the specifications and properties of the prepared samples (S1-S5) and the percentage of total volatile oils in each sample. These results also showed that geographical origins significantly affected the percentage of total VOCs influenced by environmental factors and cultivation conditions.

Table 1. Plant materials and their specifications.

Group of Samples	Origins		Growth conditions	Drying methods	latitude	longitude	Average VOCs (%)
S1	Hamedan	Gashani	Naturally grown (AP)	Shade	34°-35'-59.857"	48°-34'-38.988"	1.97
S2		Arzanfood			34°-39'-26.74"	48°-25'-52.724"	2.09
S3		Pisteski			34°-43'-7.115"	48°-26'-13.784"	1.82
S4	Tehran		On-farm cultivated (AP)		35.6892° N	51.3890° E	0.28
S5	Hamedan				34.7608° N	48.3988° E	0.98

AP: Aerial parts.

Hyperspectral imaging device and data collection

A portable HSI system was used to scan and collect spectral images of the samples. The system includes a portable VIS-NIR hyperspectral imager (Specim IQ, Specim, Spectral Imaging Ltd) that covers the spectral range from 400 to 1000 nm at 3 nm increments and 204 spectral bands with spatial data resolution of 512×512 pixels (hypercube dimensions: 512×512×204). The system also includes two halogen-based illumination sources recommended by Specim Co. that cover the full 400 to 1000 nm range for smooth lighting of the field of view surface of the camera. It could be used as a table-based device in the laboratory or as a portable screening system in stores or markets. Figure 1 shows the schematic representation of the developed HSI system, spectra hypercube, and extracted spectra. The plant samples were ground into small parts, sieved with a 2 mm sieve size, and put in petri dishes (5 cm in diameter and 2 cm in height). To prevent light scattering, the samples were pressed to be flatted before scanning. The spectral reflectance of 12 samples was recorded at the same time using the camera recording Software. Then, a three-dimensional (3-D) hypercube containing 204 image layers was obtained.

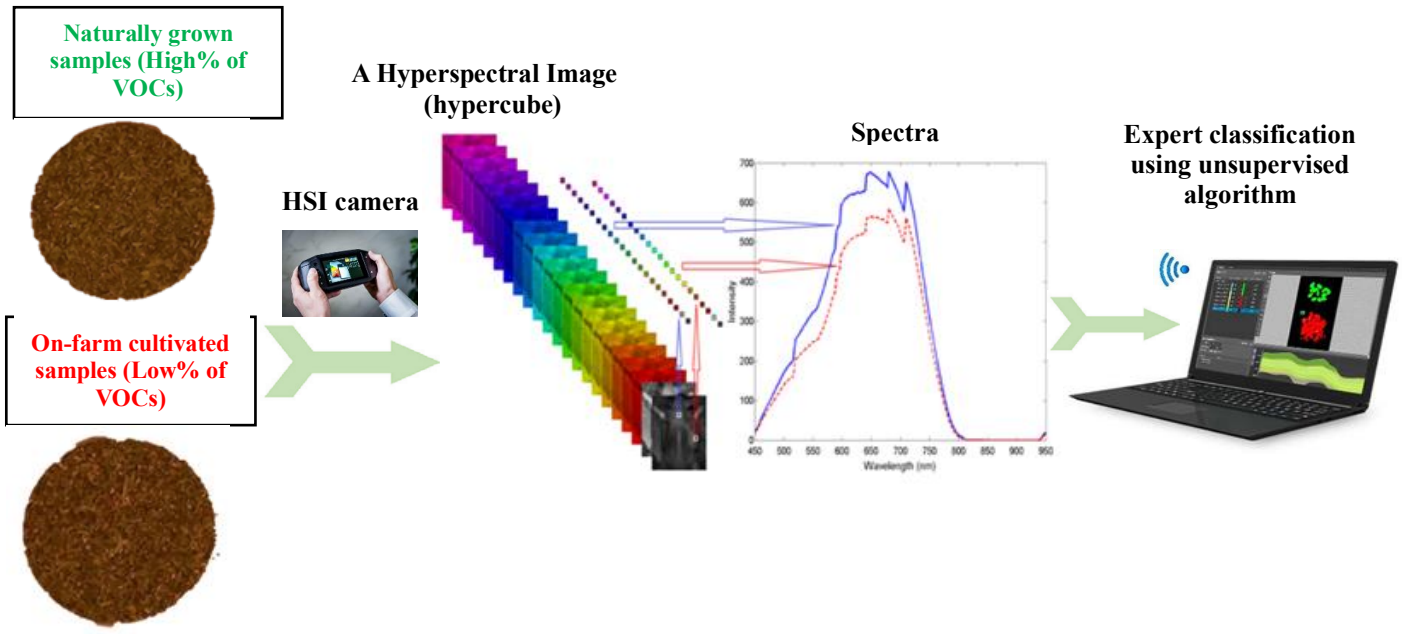

Fig. 1. Schematic representation of the intelligent HSI system developed for detecting non-authenticity in *N. crispa*.

Before scanning the target area, the camera recorded the black-and-white references and automatically makes the reflectance calculation for each measurement. Images were captured and saved to the camera's SD card and then transferred to the laptop computer. The region of interest (ROI) of each sample (30×30 pixels) was selected, and then the spectra (900 spectra of 900 pixels for each sample) were saved. The capturing was repeated for five samples from each origin/farm, and then the averaged spectra of the samples were recorded for further data analyses. Figure 2 shows the preprocessed (using SG filter) and averaged spectral reflectance curves of the *N. crispa* samples, respectively.

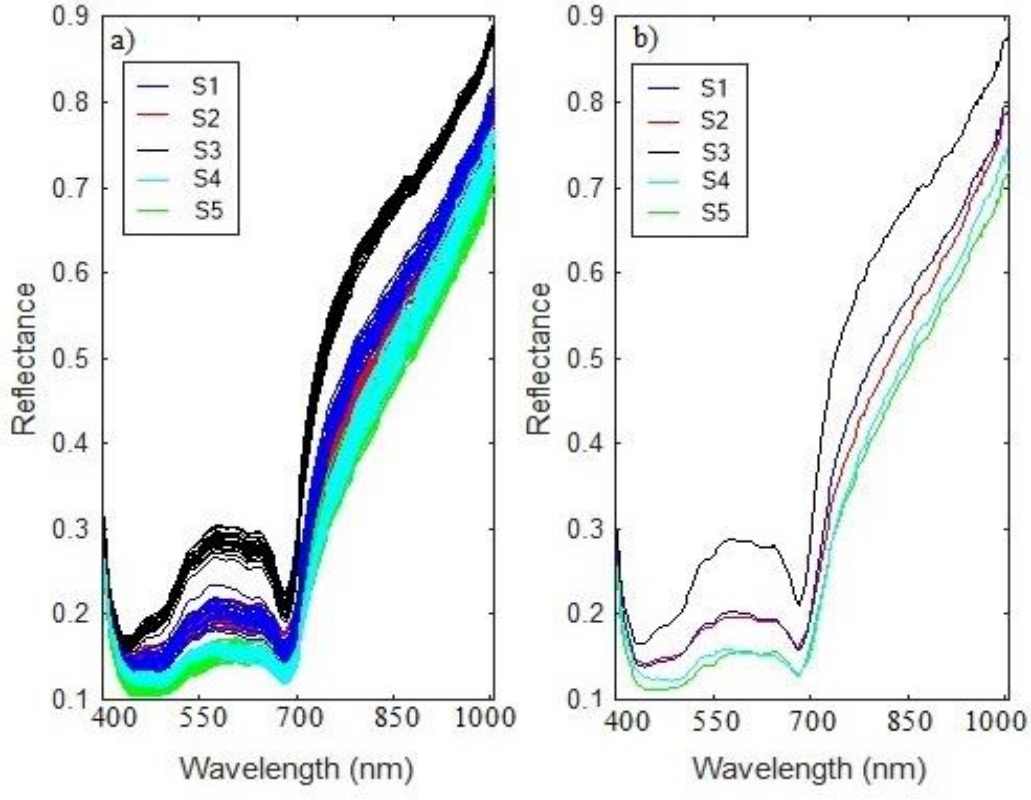

Fig. 2. (**a**) Preprocessed spectra of the *N. crispa* samples from each origin/farm (S1-S5, blue, red, black, cyan, and green, respectively); and (**b**) averaged spectrum of the samples.

The spectral curves of the *N. crispa* samples were better distinguished by wavelengths in the spectral band of 550-650 and 750-900 nm, although overlaps still exist. The wavelengths related to N-H, C-H, and O-H chemicals and associated with the oil contents in the food products are in the range of 800-900 nm (Kamruzzaman et al., 2016). The wavelengths in the spectral band of 920-980 nm are related to O-H due to the moisture content in the samples (Wu et al., 2012). To reduce the spectra dimensions, PCA was applied to project the data onto a three-dimensional PC space (PC1, PC2, and PC3). It helps to make more effective and high-dimensional addressable data and to reduce the number of input data points for further analysis. Next, two bio-inspired unsupervised algorithms were created to cluster the *N. crispa* samples' spectral fingerprints. Crisp clustering was applied using the SOM neural network and automatic clustering was performed using the ABC algorithm.

Crisp clustering using SOM

The SOM is a machine-learning algorithm that groups data according to their similarity. It was introduced by Teuvo Kohonen in the early 80s (Kohonen, T. 1982a). To reveal the input data distribution, it uses a set of finite models that are routinely related to the nodes of a regular grid so that more similar models connect to its neighborhood nodes and less similar ones keep farther apart in the grid. This strategy presents an intelligence into the topographic relationships of high-dimensional data items. The created grid using the SOM can be calibrated as an adjustable net according to the information that belongs to certain predetermined classes. It is used as an appropriate projection model for illustrating the inherent grouping structure of

unknown input data (Kohonen, 2013). It has also been extensively applied for data exploration in different fields of research such as physical sciences, industries, and finance (Cottrell et al., 2018). Many popular performance evaluation methods have been suggested to evaluate clustering models, but many of them do not work well in complicated conditions when clusters are presented with different densities. Chou et al. (2004) proposed a new index, the CS index, which is a criterion of within-cluster similarity to between-cluster dissimilarity according to Equation 1.

$$S = \frac{\sum_{i=1}^{k}\left(\frac{1}{ni}\sum_{X_i \in C_i} \max_{X_u \in C_i}\{\|X_i - X_u\|\}\right)}{\sum_{i=1}^{k}\left(\min_{j \in k, j \neq i}\{\|m_i - m_j\|\}\right)} \quad (1)$$

In which k and n are the number of clusters and number of samples in a cluster, respectively, m indicates the cluster centroid, and Xi is the pattern in cluster Ci. A smaller CS index value in a clustering structure would be an appropriate grouping.

Automatic clustering using ABC

In crisp clustering, a challenging problem is to set up the number of clusters. This is well-known as the clustering problem and is more critical when the input data has many dimensions of variation, especially when overlapping exists among clusters, and when clusters differ significantly in shape, size, and density. Unsupervised automatic clustering was introduced to overcome this inconvenience (Jose Garcia and Gomez-Flores, 2016; Rahimzadeh et al., 2022). Since prior domain knowledge of the data is not provided, automatic clustering is used to assign an ideal number of clusters in the datasets. In this study to do the automatic clustering of the *N. crispa* samples spectra, a coding plan was designed to change the number of clusters within the range of [K_{min}, K_{max}]. It plays a pertinent role in the efficacy of any meta-heuristic and organizes an indispensable step in its design. The encoding scheme was performed from the lowest (2) to the highest probable clusters (10), K_{min}= 2 to K_{max}= 10. Each solution shows that the center of the cluster was encoded into a Kmax × (T+1) vector within the range of [0, 1]. T is the three-dimensional PC space. The activation thresholds are determined by the inputs of the (T+1)-th column in each solution vector (Talbi, 2016). They are used to specify which clusters are engaged in the clustering process. The encoding efficiency depends on the utilized search operators such as neighborhood, mutation, and recombination. The ABC algorithm, recently introduced as a swarm-based algorithm, imitates the intelligent foraging behavior of a honeybee swarm (Karaboga and Basturk, 2009). It was used to optimize the automatic clustering of the *N. crispa* spectra. In this method, the environment of a hive is simulated as the search space. The ABC algorithm starts with bees that randomly select several food sources (solution vector (x)) and evaluate their nectar amount (fitness of the solution vector) for the clustering scheme (Davies and Bouldin, 1979). The Davies-Bouldin (DB) index is used to calculate the fitness of the solution vector. This index indicates the within-cluster variance ratio to between-cluster distance. The xij represented food sources where i={1, …., number of the initial sources} and j={1, …., number of the parameters to be optimized}. Then the previously discovered areas (xij) are operated by the bees to make new food sources (vij). The process is done according to Equation 2.

$$v_{ij} = x_{ij} + \emptyset_{ij}(x_{ij} - x_{bj}) \quad (2)$$

In which b and j are randomly chosen index and variable, respectively, and ϕ_{ij} is a real random digit within the range [-1,1]. According to this equation, the information from the vij sources is more beneficial than the previously discovered sources shared by the bees to the onlooker bees. The newly trained bees utilize the information and try to find a new beneficial food source. In each turn, the sources that did not have enough potential to be better are supposed to be desolated sources and their relevant bees become scouts to look for new sources in the area. The initial clusters go through the successive processes of the onlooker and scout bees for a predefined set of turns. Eventually, the optimum clustering structure is introduced by the best source. The performance of the ABC algorithm is better than or similar in comparison with other meta-heuristic algorithms such as the genetic algorithm, particle swarm optimization algorithm, differential evolution algorithm, and evolution strategies. It has the advantages of employing fewer control parameters, high flexibility, robustness against the initialization step, and fast convergence speed (Akay and Karaboga, 2012). More detailed discussions of the performance of the ABC algorithm in comparison with other population-based algorithms in different optimization problems are reported by Karaboga and Akay (2009).

The system developed in this study is a one-class modeling classifier. In machine learning, one-class classification (OCC), also known as unary classification or class modeling, aims to identify objects of a specific class among all objects by primarily learning from a training set that contains only the objects of that class. This approach differs from the traditional classification problem, which seeks to distinguish between two or more classes using a training set that includes objects from all the classes. One-class classification is generally more challenging.

Results

Original grouping using PCA

The PCA scores plot was utilized for the original clustering of the *N.Crispe* spectra (Fig 3a). It normally happens according to the sample spectral properties as a map of PC1 plotted against PC2. The relative importance of each PC is expressed in terms of how much variance of the original data is described. It is also observed how the 5 *N. crispa* samples (S1-S5) are distributed along the PC1 (94%) axis. Adjacent samples with close scores along a PC have almost equal values for the corresponding variables and are similar. On the contrary, samples for which the value of their score varies significantly (S4 and S3) are exactly different from each other. Sample 4 is distinguished, whereas the others are overlapped, and the problem was that sample 5 was not distinguished from S1-S3 (natural samples). Figure 3b shows the PCA loading plot, which is commonly used in spectral data interpretation and describes the relationships between variables. It also indicates the actual dimensions of the data transformed by PCA.

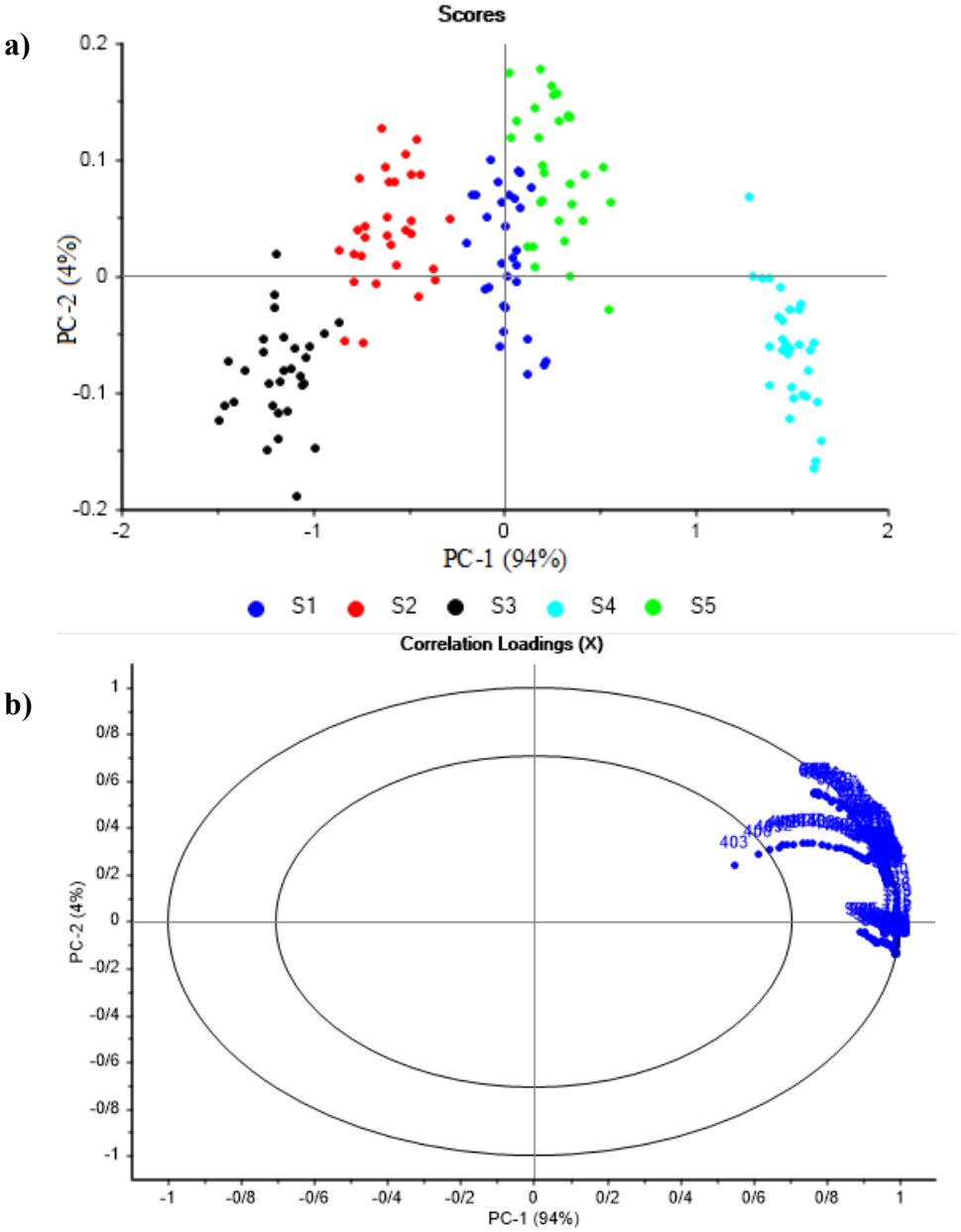

Fig. 3. (**a**) PCA scores plot of *N. crispa* samples (400-1000 nm) and; (**b**) PCA loading plot of the whole variables (wavelengths).

Variables that lie close together are highly correlated, and for each PC, variables with high loadings (i.e., close to +1 or –1), are more effective in the sample distinguishing. This interpretation is a

key direction and will help to decide how many data dimensions will be taken and excluded (dimension reduction or noise cancellation) for further analysis. The PCA automatically excluded 55 spectra (bands of 400-457, 670-721, 952-1000 nm) and reduced the 204 spectra to 149 spectra. The remainder of the spectra, including 800-900 nm, are associated with the oil content in *N. crispa*. This issue was already proven by Kamruzzaman et al. (2016).

Crisp clustering by SOM

Since the SOM algorithm needs a pre-set number of clusters, it (the input neurons) was set equal to the number of the *N. crispa* samples. Therefore, a linear grid, 5×1, was defined for each node to be correlated to each cluster center. During the calibration of SOM, the spectral characteristics of the samples were correlated to the linear grid and then the undefined distance matrix (U-matrix) between the adjacent neurons was created. The clustering results and U-matrix of *N. crispa* samples are shown in Figure 4. By comparing this result against the original grouping by the PCA (Fig. 3a), it indicated that the samples were separated in their original groups, two clusters, natural and on-farm cultivated samples, but there were intra-group confusions between the clusters S5 & S4 and S1, S2, & S3.

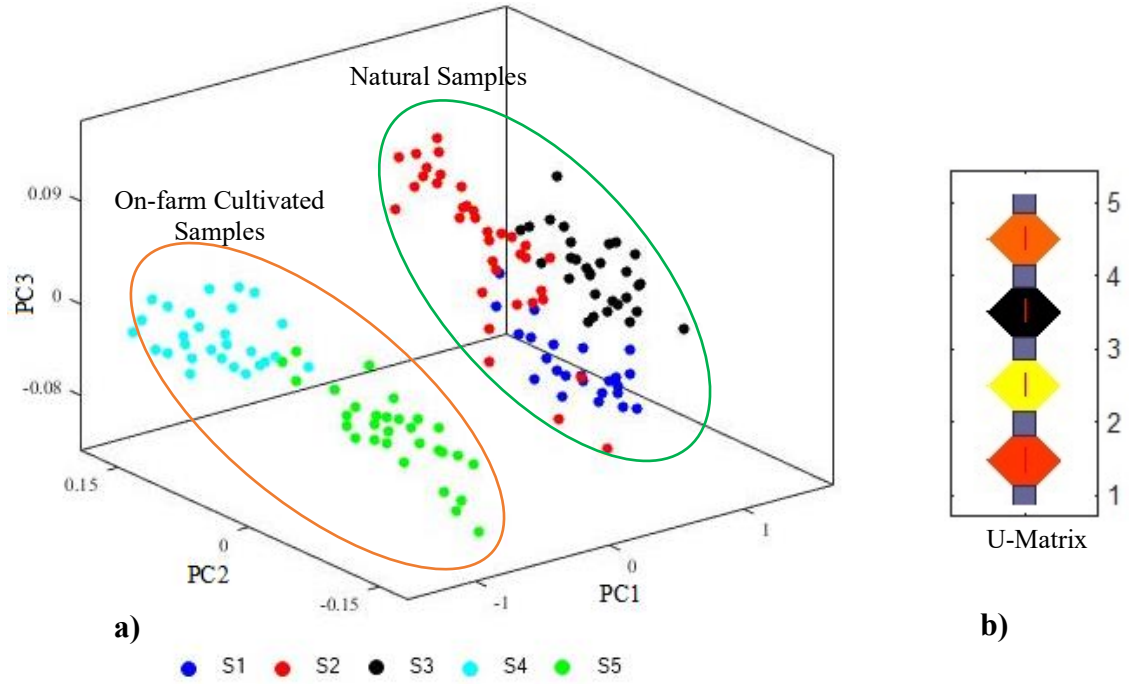

Fig. 4. (**a**) Crisp clustering structure of the natural and on-farm cultivated samples using the SOM algorithm and; (**b**) U-matrix showing the distance between the samples, darker colors represent greater distances, while lighter colors represent shorter distances.

U-Matrices with color scales or indicative coloring are used to compare the distance between the neighboring cluster centers. Dark coloring between the samples means the samples become farther, and on the contrary, light coloring between the samples means a proportionally short distance between them. For the *N. crispa* samples, the darker color occurred between samples S1, S2, & S3 and S4 & S5, indicating the longest distance between them. Red and orange colors relatively show a shorter distance between S1 & S2 and S4 & S5, respectively. The brightest color was between clusters S2 & S3, indicating the shorter distance between them. By comparing and relating the *N. crispa* samples spectra and their total volatile oils content (Table 1) to the coloring patterns, it can be concluded that because the natural samples have a higher percentage of VOCs and surly different spectra, their cluster gets a further distance from the samples S4 and S5 which had a lower percentage of VOCs. The value of the CS index and its components (between-group separation and within-group scatter) for the *N.Crisps* samples is given in Table 2. In the report of this measure, the distance between the clusters in the *N.Crisps* samples was more than twice as big as the distance within the cluster scatters (Numerator). As was already stated in using the U-matrices, the discrepancy between the two main clusters was due to the drastic variation in their spectral characteristics, which can be interpreted as the result of the big difference in their VOCs.

Table 2. Values of the CS index obtained in the clustering of *N. crispa* using Crisp Clustering.

Parameter	Value
CS index	0.4896
Within cluster scatter (Numerator)	1.9776
Between cluster separation (Denominator)	4.0394

Since the CS index is a cluster validation method for unsupervised algorithms, it was used as a validation method to assess how well the classes are separated from one another. The information provided in the text expresses the separation of the groups from each other as well as the compactness of the groups.

Automatic clustering using ABC

The real clustering of the samples' spectra included two main clusters concerning their growing condition. Since the number of samples was not predefined in the automatic clustering, the ABC was created in several performances. Results of the clustering operation using the automatic clustering algorithm are depicted in Figure 5. As the algorithm began to cluster the samples into two groups (Fig. 5a), the members of the closest clusters (S1-S3 and S4 & S5) were considered integrated clusters, and the value of the DB index was calculated at 0.35961. The lower the DB value, the more appropriate clustering would be. These two main clusters had the longest distance, and the minimum DB index value was also obtained. Therefore, the automatic clustering using ABC could detect these two main groups, which was the main goal of this study. When the samples were clustered into 3 clusters, the members of S1 & S2, S3, and S4 & S5 were grouped in 3 clusters separately (Fig. 5b). For this structure, the S1 & S2 and S4 & S5 shared their members to form individual clusters and made single clusters, respectively. When the samples were clustered into 4 clusters, S1 & S2, S3, S4, and S5 were grouped into four clusters (Fig. 5c). In this structure, S1 & S2 shared their members. Moreover, when the 5-cluster was set, the algorithm clustered all the samples in 5 groups (Fig. 5d).

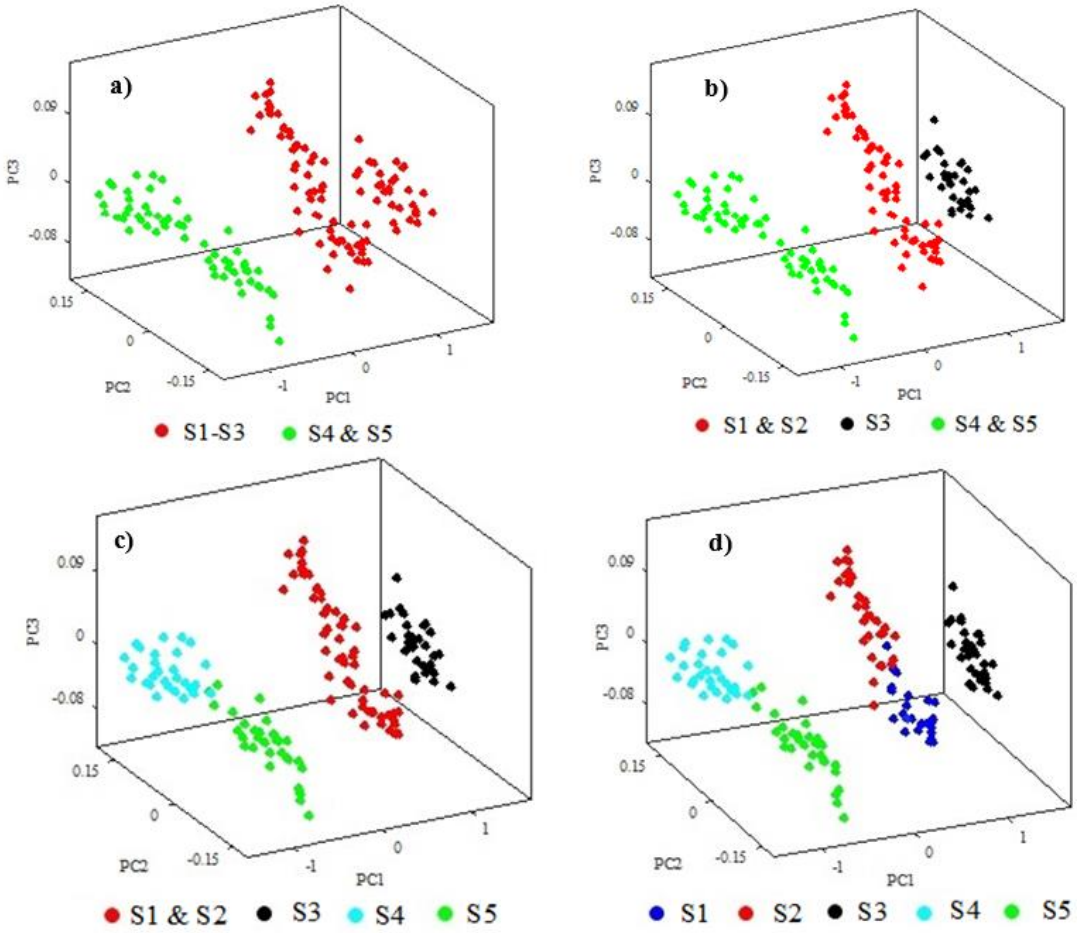

Fig. 5. Automatic clustering of *N. crispa* samples into (**a**) two clusters, in which the natural samples were segregated from the on-farm cultivated samples; (**b**) three clusters; (**c**) four clusters; and (**d**) five clusters.

Discussion

A lower DB index value signifies superior clustering quality, as it indicates clusters that are compact and well-separated from one another. The DB index is calculated by taking the average 'worst-case' score for each cluster, where the score for a cluster is the maximum ratio of within-cluster scatter to between-cluster separation for that cluster and any other cluster. All in all, when automatic clustering was applied, the number of gained clusters was similar to the original grouping but more separable. This is specific because this approach plays smartly and attempts to combine the samples with similar spectra. It was also found that the automatic clustering could intelligently recognize the *N. crispa* samples that do not have enough differences as individual clusters. It is crisp in setting apart the samples into groups but flexible in the number of clusters. Automatic clustering attempts to make a structure that includes several clusters with the most similarity of the data within the clusters. It tries to reach the criteria of intra-cluster similarity to between-cluster separability. In cases where the pre-set number of clusters varies from the original clusters of the input data structure, the automatic clustering by ABC algorithm eliminates the limitation of the pre-set number of clusters with high probability and is capable of making a structure with the clusters according to the spectra changes of the samples from different geographical origins.

In previous research regarding the application of the HSI coupled with unsupervised machine learning algorithms, Mayatopani et al. (2023) developed an HSI+SOM system to classify weed leaves with different medicinal properties. They found an accuracy of 89.44% in grouping their samples. Izadi and kiani (2024) utilized an HIS + automatic clustering algorithm by the ABC to differentiate various types of pomegranate molasses samples and evaluate the possibility of date syrup adulteration detection in the samples. The physicochemical properties of the samples (brix index, sucrose, acidity, ash content, pH, and formalin index) were measured as the reference data. They illustrated that the developed system could detect date syrup nonauthenticity in pomegranate molasses samples from the level of 5% adulteration.

Conclusions

A portable HSI was coupled with two biologically inspired algorithms, SOM and automatic clustering by ABC, for unsupervised authenticity analysis of *N. crispa* samples. In real applications, the system usually faces a variety of unknown samples from different geographical origins, which are challenging conditions for the developed supervised algorithm. In performing the SOM clustering, it was generally found that the two main groups, naturally grown and on-farm cultivated samples, were distinguished into their pertinent groups. The U-matrix was used to describe the amount of spectral change in the *N. crispa* samples. The spectra of the naturally grown and on-farm cultivated samples were distinguished by a significant distance. It indicated the variation of their VOCs in their growing conditions. In some cases, the clusters overlapped with each other, indicating the similarity of their spectral fingerprints. However, the automatic clustering by the ABC algorithm did not need a predefined number of clusters but successfully detected the spectral fingerprints of the naturally grown (S1-S3) and on-farm cultivated samples (S4 & S5) from each other as well as their samples in each original group. The automatic clustering by the ABC presented a smarter performance to distinguish real intrinsic clusters with the most within-group similarity, in comparison with crisp clustering using SOM. The overall results demonstrated the great potential of the HSI system coupled with biologically inspired unsupervised automatic algorithms as a reliable screening tool for rapid, nondestructive, and real-time authenticity evaluation of *N. crispa* samples in the markets. The proposed HSI system could also be calibrated and used for monitoring and authenticity evaluation of other types of medicinal and aromatic plant products.

Funding

This project was supported by Sari Agricultural Sciences and Natural Resources University (grant number 02-1400-02). We gratefully acknowledge the financial support from the above source.

Conflict of Interest

The authors indicate no conflict of interest in this work.

References

Acierno, V., Fasciani, G., Kiani, S., Caligiani, A., van Ruth, S. 2019. PTR-QiToF-MS and HSI for the characterization of fermented cocoa beans from different origins. *Food Chemistry*, 289, 591–602. DOI: 10.1016/j.foodchem.2019.03.095

Akay, B., & Karaboga, D. 2012. A modified artificial bee colony algorithm for real-parameter optimization. Information Sciences, 192, 120e142. DOI: 10.1016/j.ins.2010.07.01

Chou, C. H., Su, M. C., & Lai, E. 2004. A new cluster validity measure and its application to image compression. Pattern Analysis & Applications, 7(2), 205-220. DOI: 10.1007/s10044-004-0218-1

Cottrell, M., Olteanu, M., Rossi, F., Villa-Vialaneix, N. 2018. Self-Organizing Maps, theory, and applications. Revista de Investigacion Operacional, 39 (1), 1-22.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 1(2), 224-227. DOI: 10.1109/TPAMI.1979.4766909

Edris, M., Ghasemi-Varnamkhasti, M., Kiani, S., Yazdanpanah, H., & Izadi, Z. 2024. Identifying the authenticity and geographical origin of rice by analyzing hyperspectral images using unsupervised clustering algorithms. Journal of Food Composition and Analysis, 125, 105737. DOI: 10.1016/j.jfca.2023.105737

Edris, M., Ghasemi-Varnamkhasti, M., Kiani, S., Yazdanpanah, H., Izadi, Z. 2025. Application of fuzzy clustering algorithm and hyperspectral images for rice authentication. Iranian Journal of Biosystems Engineering, 56, 1-15. DOI: 10.22059/ijbse.2025.384493.665571

Eshkabilov, S., Lee, A., Sun, X., Lee, C.W., Simsek, H. 2021. Hyperspectral imaging techniques for rapid detection of the nutrient content of hydroponically grown lettuce cultivars. Computers and Electronics in Agriculture, 181, 105968. DOI: 10.1016/j.compag.2020.105968

Hu, N., Li, W., Du, C., Zhang, Z., Gao, Y., Sun, Z., Yang, L., Yu, K., Zhang, Y., Wang, Z. 2021. Predicting micronutrients of wheat using hyperspectral imaging. Food Chemistry, 343, 128473. DOI: 10.1016/j.foodchem.2020.128473

Izadi, Z., & Kiani, S. 2024. Pomegranate molasses authentication using hyperspectral imaging system coupled with automatic clustering algorithm. Journal of Food Science, 1–13. DOI: 10.1111/1750-3841.17134

Jose-Garcıa, A., & Gomez-Flores, W. 2016. Automatic clustering using nature-inspired metaheuristics: A survey. Applied Soft Computing Journal, 41, 192-213. DOI: 10.1016/j.asoc.2015.12.001

Karaboga, D., & Akay, B. 2009. A comparative study of Artificial Bee Colony algorithm. Applied Mathematics and Computation, 214(1), 108-132. DOI: 10.1016/j.amc.2009.03.090

Karami, M., Abadi, M.T., Ayyari, M. 2021. Evaluation of changes in the quantity and quality of essential oil of *Nepeta Crispa* Willd in different natural crop habitats. Eco-Phytochemical Journal of Medicinal Plants, 2, 1-12. (In Persian)

Kamruzzaman, M., Makino, Y., Oshita, S. 2016. Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging. Food Chemistry, 196, 1084–1091. DOI: 10.1016/j.foodchem.2015.10.051

Ke, J., Rao, L., Zhou, L., Chen, X., Zhang, Z. 2020. Non-destructive determination of volatile oil and moisture content and discrimination of geographical origins of Zanthoxylum bungeanum Maxim. By hyperspectral imaging. Infrared Physics and Technology, 105, 103185. DOI: 10.1016/j.infrared.2020.103185

Kiani, S., Van Ruth, S. M., and Minaei, S. 2018. Hyperspectral imaging, a non-destructive technique in medicinal and aromatic plant products industry: Current status and potential future applications. Computers and Electronics in Agriculture, 152, 9–18. DOI: 10.1016/j.compag.2018.06.025

Kiani, S., Van Ruth, S.M., Minaei, S. 2019. Hyperspectral Imaging as a Novel System for Nutmeg Authenticity Evaluation. LWT, Food Science and Technology, 104, 61-69. DOI: 10.1016/j.lwt.2019.01.045

Kiani, S., Yazdanpanah, H., Feizy, J. 2023. Geographical origin differentiation and quality determination of saffron using a portable Hyperspectral imaging system. Infrared Physics & Technology, 131, 104634. DOI: 10.1016/j.infrared.2023.104634

Kiani, S., Rahimzadeh, H., Kalantari, D., Moradi-Sadr, J. (2023). Aroma modeling and quality evaluation of spearmint (Mentha spicata subsp. spicata) using electronic nose technology coupled with artificial intelligence algorithms. Journal of Applied Research on Medicinal and Aromatic Plants, 35, 100473. DOI: 10.1016/j.jarmap.2023.10047

Kohonen, T. 1982a. Analysis of a simple self-organizing process. Biol. Cybern, 44:135–140. DOI: 10.1007/BF00317973

Kohonen, T. 2013. Essentials of the self-organizing map. Neural Networks, 37, 52-65. https://doi.org/10.1016/j.neunet.2012.09.018.

Liu, F., Wang, F., Liao, G., Lu, X., Yang, J. 2021. Prediction of Oleic Acid Content of Rapeseed Using Hyperspectral Technique. *Applied Sciences*, 11, 5726. DOI: 10.3390/app11125726

Mayatopani, H., Handayani, N., Sabti Septarini, R., Nuraini, R., Heriyani, N. 2023. Implementation of Self-Organizing Map (SOM) Algorithm for Image Classification of Medicinal Weeds. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 7(3), 437 – 444. DOI: 10.29207/resti.v7i3.4755

Mojab, F., Nickavara,B., Hooshdar Tehrani, H. 2009. Essential Oil Analysis of Nepeta Crispa and N. Menthoides from Iran. Iranian Journal of Pharmaceutical Sciences, 5(1), 43-46. DOI:10.22037/ijps.v5.41152

Nejadhabibvash, F., Rezaee, M. B., Mahmudi, A., Jaimand, K. 2018. Effect of Harvesting Time on Content and Chemical Composition of Essential Oil from Stachys lavandulifolia Vahl (Lamiaceae). Journal of Medicinal plants and By-product, 7(2), 181-187.

Odabas, M.S., Simsek, H., Lee, C.W., Iseri, I. 2017. Multilayer perceptron neural network approach to estimate chlorophyll concentration index of lettuce (*Lactuca sativa* L.). Commun. Communications in Soil Science and Plant Analysis, 48 (2), 162–169. DOI: 10.22092/jmpb.2018.118146

Orrilloa, I., Cruz-Tiradob, J.P., Cardenasa, A., Orunaa, M., Carnero, A., Barbinb, D.F., Sichea, R. 2019. Hyperspectral imaging as a powerful tool for identification of papaya seeds in black pepper. Food Control, 101, 45–52. DOI: 10.1016/j.foodcont.2019.02.036

Petrakis, E. A., Cagliani, L.R., Tarantilis, P. A., Polissioua, M.G., Consonni, R. 2017. Sudan dyes in adulterated saffron (*Crocus sativus L.*): identification and quantification by 1 H NMR. Food Chemistry, 217, 418–424. DOI: 10.1016/j.foodchem.2016.08.078

Qin, J., Chao, K., Kim, M.S., Lu, R., Burks, T.F. 2013. Hyperspectral and multispectral imaging for evaluating food safety and quality. Journal of Food Engineering, 157–171. DOI: 10.1016/j.jfoodeng.2013.04.001

Rahimzadeh, H., Sadeghi, m., Mireei, S.A., Ghasemi-Varnamkhasti, M. 2020. Unsupervised modelling of rice aroma change during ageing based on electronic nose coupled with bio-inspired algorithms. Biosystems Engineering, 216, 132-146. DOI: 10.1016/j.biosystemseng.2022.02.010

Rehman, N.U., Hussain, J., Ali, S., Hussain, H., Abbas, G., Bakht, N., Al-Harrasi, A. 2020. Chemical constituents of the essential oil of Nepeta distans. Chemistry of Natural Compounds, 56(1): 159-160. DOI: 10.1007/s10600-020-02973-9

Sefidkon, F., Jamzad, Z. and Mirza, M. 2006. Chemical composition of the essential oil of five Iranian Nepeta species (*N. crispa*, *N. mahanensis*, *N. ispahanica*, *N. eremophila* and *N. rivularis*). Flavour and Fragrance Journal, 21: 764-767. DOI: 10.1002/ffj.1668

Sytar, O., Brestic, M., Zivcak, M., Olsovska, K., Kovar, M., Shao, H., He, X. 2017. Applying hyperspectral imaging to explore natural plant diversity towards improving salt stress tolerance. Science of the Total Environment, 578, 90–99. DOI: 10.1016/j.scitotenv.2016.08.014

Talbi, E.G. (2009). Metaheuristics from Design to Implementation. John Wiley and Sons, 2009. DOI:10.1002/9780470496916

Van Haute, Sam., Nikkhah, A., Malavi, D., Kiani, S. 2023. Prediction of essential oil content in spearmint (Mentha spicata) via near-infrared hyperspectral imaging and chemometrics. Scientific Reports, 13(1), 4261. DOI: 10.1038/s41598-023-36583-6

Wu, D., Shi, H., Wang, S., He, Y., Bao, Y., Liu, K. 2012. Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system. Analytica Chimica Acta, 726, 57–66. DOI: 10.1016/j.aca.2012.03.038