Research Journal of Pharmacognosy (RJP) 2(3), 2015: 1-8
Received: Mar 2015
Accepted: May 2015

Original article

ATR-IR fingerprinting as a powerful method for identification of traditional medicine samples: a report of 20 herbal patterns

A. Sakhteman[1], R. Keshavarz[2], A. Mohagheghzadeh[2,3*]

[1]*Department of Medicinal Chemistry, Shiraz University of Medical Sciences, Shiraz, Iran.*
[2]*Department of Phytopharmaceuticals (Traditional Pharmacy), Faculty of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran.*
[3]*Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences, Shiraz, Iran.*

Abstract

Background and objectives: Attenuated total reflectance-inferared (ATR-IR) spectra can be used as a non-invasive fingerprinting approach in quality control of herbal samples. **Methods**: Twenty versatile herbal samples were subjected to attenuated total reflectance-inferared (ATR-IR) spectroscopy followed by different clustering methods in order to determine by which method more reasonable classifications would be obtained. **Results:** All classification methods (K-means, HCA, PCA and SOM) were able to discriminate the two medicinal seeds, *Hyocyamus niger* and *Peganum harmala* from other herbal samples. Similarly, the starch samples were clustered in a reasonable method. In HCA, one cluster included three types of starch samples: *Zea mays*, *Oryza sativa* and *Triticum aestivum*. All the four classification methods were able to separate *Solanum tuberosum* starch from other starch samples. HCA and SOM, were able to classify leaf samples *Origanum vulgare* and *Melissa officinalis* belonging to Lamiaceae family, in one category. *Crocus sativus* and its adulterant *Carthamus tinctorius* flowers were identified by PCA, HCA and SOM as different categories. **Conclusion:** The result of this study can be utilized for identification and quality control of traditionally used medicinal plant samples in an unknown sample powder. Such data could be the basis for preparing a data bank on Iranian medicinal samples which in turn is used as a simple, fast and reliable method for characterization of herbal powders in Pharmacopoeias.

Keywords: ATR-IR, fingerprint, medicinal plants, traditional medicine

Introduction

The fast increasing trend of science and technology in the field of traditional medicine (TM) and phytopharmaceuticals makes it inevitable not to extend approving criteria for herbal products [1]. Regarding the importance of designing more sensitive and applicable methods of quality control for herbal complex and compound samples normally used in TM, some novel applications of instrumental techniques such as infrared (IR) spectroscopy have been discussed in recent years [1]. Attenuated total reflectance (ATR) is a novel non-invasive method with less time consumption property compared to present chromatography or

*Corresponding author: mohaghegh@sums.ac.ir, Tel: +9871-32426070, Fax: +9871-32424456

microscopic methods [1,2]. Another feature of this technique is its reproducibility and ease of use for different liquid and solid samples. In order to make this technique responsive and applicable in quality control, classification or clustering methods should be coupled with experimental IR data. In classification, a set of predefined classes are predefined while clustering tends to groups a set of objects and find the possible relationship between the objects. Based on the context in machine learning, classification is supervised learning and clustering is unsupervised learning [3]. Supervised learning are used in cases, where the training data are defined as the examples of the input vectors along with their corresponding target vectors. In pattern recognition problems, the training data include a set of input vectors without any corresponding target values. The goal in such unsupervised learning problems is to discover groups of similar examples within the data [3]. For example in one study Roasted and raw Semen Cassiae have been classified by IR fingerprinting approach [4]. Identification of adulterations such as sibutramine in supplements by ATR-IR and chemometrics methods has been also reported [5]. The effect of heat processing methods on the stability of the canola seed proteins using ATR-FT/IR spectroscopy has been discussed [6].

In microscopic analysis of the samples, the diagnosis is based on finding certain organelle or parts of the herbs and comparing them with the available atlas of the medicinal plants [7-9]. Such pharmacopeia's issues are therefore very much dependent to the observer claims about that sample, based on macroscopic appearance, organoleptic characters and microscopic characteristics [8,10]. On the other hand, spectroscopic methods are relied upon the unique pattern observed for each herbal sample [7,8]. The unique pattern of each sample could be the reflection of primary as well as secondary metabolites presented in those natural materials. Another application of this technique could be in preparation of standardized finish products using calibration curves. The fingerprint approach for quality assessment has been approved by the Chinese State Food and Drug Administration [4]. Both supervised and unsupervised techniques have been applied for discriminative identification and classification purposes [4]. The most used methods for classification included principal component analysis (PCA), K-nearest neighbors (KNN), linear discrimination analysis (LDA), cluster analysis and soft model of class analogies (SIMCA). In this study 20 herbal samples of Iranian Traditional Medicine (ITM) were subjected to IR spectroscopy using ATR. The resulted data were subsequently clustered by unsupervised classification methods including PCA, HCA, K-means and Kohonen neural network.

Experimental

Plant materials

Twenty herbal materials were randomly selected from those presented in medicinal plants laboratory for training pharmacy students. The selected plant samples are as it is described in table 1: leaves (*Hyoscyamus niger*, *Aloysia citriodora*, *Digitalis nervosa*, *Urtica pilulifera*, *Origanum vulgare*, *Melissa officinalis*, *Myrtus communis*, *Adiantum capillus-veneris*, *Atropa belladonna*, *Fumaria parviflora*, *Camellia sinensis*); root (*Glycyrrhiza glabra* var. *glabra*); starches (*Zea mays*, *Solanum tuberosum*, *Oryza sativa*, *Triticum aestivum*); flowers (*Carthamus tinctorius*, *Crocus sativus*); gum (*Hymenaea verrucosa*) and seed (*Peganum harmala*).

Plant sample specimens (PM) were authenticated by S. Khademian and a sample was deposited at the Herbarium of Faculty of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran.

ATR-IR spectroscopy and data pretreatment

In order to investigate the similarities between different samples, IR spectroscopy was used to obtain the fingerprint pattern of all cases. For this purpose, different powders were subjected to a Bruker vertex-70 instrument using ATR

apparatus. Before acquisition, the baseline for all data was corrected in order to suppress the ingredients in the chamber. The transmittance values for all samples were obtained in the middle IR range 600-3400 cm^{-1}. The resulted data were subsequently entered into Matlab software for further classification analysis. In order to suppress the baseline fluctuations, Standard Normal Variety (SNV) was performed as a preprocessing technique on the data matrix.

K-means analysis

K-means partitions the points in the N-by-P data matrix into K clusters. This partition minimizes the sum of the within-cluster sums of point-to-cluster-centroid distances. Rows of X correspond to points, columns correspond to variables. K-means returns an N-by-1 vector containing the cluster indices of each point. Euclidean distances are used for clustering [11].

Hierarchical cluster analysis (HCA)

To perform HCA, the resulted matrix was subjected to MATLAB (Mathworks Inc.) software. Cluster definition was done by means of euclidean distance as a measure of similarity using unweighted pair group method (UPGMA). The plot of the distances versus samples was used to represent the data based on their similarities [12].

Principal component analysis (PCA)

To cluster the samples based on their components, PCA was used as an unsupervised clustering analysis technique. For this purpose, all the principal components (PCs) were extracted from the resulted matrix of data using singular value decomposition algorithm. PCA theory is based on ranking the PCs according to their eigenvalues in such a way that the first PC contains the most variation in the data set. Accordingly, the second PC is calculated to be orthogonal with respect to the first one. The plot of the first two PCs could represent data scattering in a two dimensional space [13].

Kohonen self-organizing maps (SOM)

Kohonen Maps (neural network analysis) as a self-organizing approach can be applied to unsupervised problem solving including cluster analysis. Similar samples are being connected to the topological vicinal neurons in the system. The weights of the neurons are based on the number of the responses in the target vectors and adapt to the locations in the ANN with similar properties towards their input vectors. Based on the input objects, the weights are updated and all the objects are introduced for a certain number of times (epochs). A Kohonen map with dimensions of 4×4, which is built for a dataset by p variables is depicted in figure 1 [14].

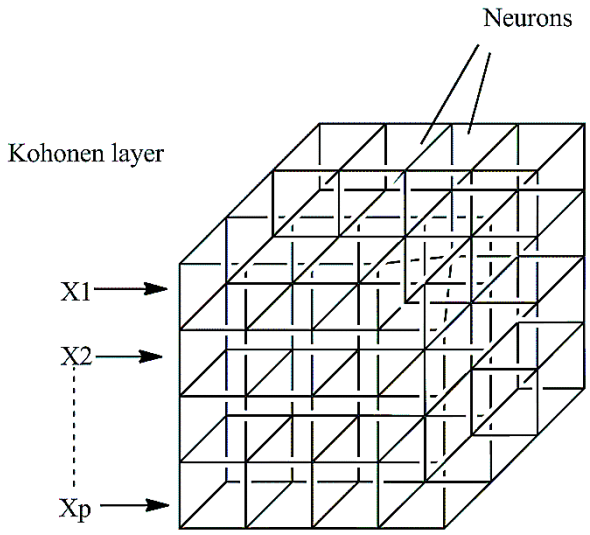

Figure 1. Kohenen maps as a self-organizing neural network analysis method [14]

Results and Discussion

Twenty versatile herbal samples (table 1) were subjected to ATR-IR spectroscopy (figure 2) followed by different clustering methods in order to determine by which method more reasonable classifications would be obtained. All the four

Table 1. The characterization of 20 plant materials subjected to ATR-IR fingerpriting

No.[1]	Type	Persian name	Scientific name	Family	PM no.[2]
1	Seed	Bazorl-ol- banj	*Hyoscyamus niger* L.	Solanaceae	792
2	Leaf	Behlimoo	*Aloysia citriodora* Palau	Verbenaceae	676
3	Leaf	Angoshtaneh	*Digitalis nervosa* Steud. & Hochst. *ex* Benth.	Scrophulariaceae	679
4	Seed	Esfand	*Peganum harmala* L.	Zygophyllaceae	683
5	Leaf	Gazane	*Urtica pilulifera* L.	Urticaceae	681
6	Root	Shirin bayan	*Glycyrrhiza glabra* L. var. *glabra*	Papilionaceae	684
7	Flower	Golrang	*Carthamus tinctorius* L.	Asteraceae	755
8	Starch	Zorrat	*Zea mays* L.	Poaceae	793
9	Leaf	Marzanjoosh	*Origanum vulgare* L.	Lamiaceae	680
10	Leaf	Badranjbooyeh	*Melissa officinalis* L.	Lamiaceae	677
11	Leaf	Moord	*Myrtus communis* L.	Myrtaceae	674
12	Leaf	Pare siavashan	*Adiantum capillus-veneris* L.	Adiantaceae	675
13	Starch	Sibzamini	*Solanum tuberosum* L.	Solanaceae	796
14	Starch	Berenj	*Oryza sativa* L.	Poaceae	794
15	Flower (stigma)	Zafaran	*Crocus sativus* L.	Iridaceae	795
16	Gum	Sandrus	*Hymenaea verrucosa* Gaertn.	Papilionaceae	803
17	Leaf	Shabizak	*Atropa belladonna* L.	Solanaceae	689
18	Leaf	Shatareh	*Fumaria parviflora* Lam.	Fumariaceae	682
19	Leaf	Chay-e sabz	*Camellia sinensis* (L.) Kuntze	Theaceae	678
20	Starch	Gandom	*Triticum aestivum* L.	Poaceae	797

[1]The plant numbers are as those indicated in table 2 and figures 3-5; [2]plant material voucher number

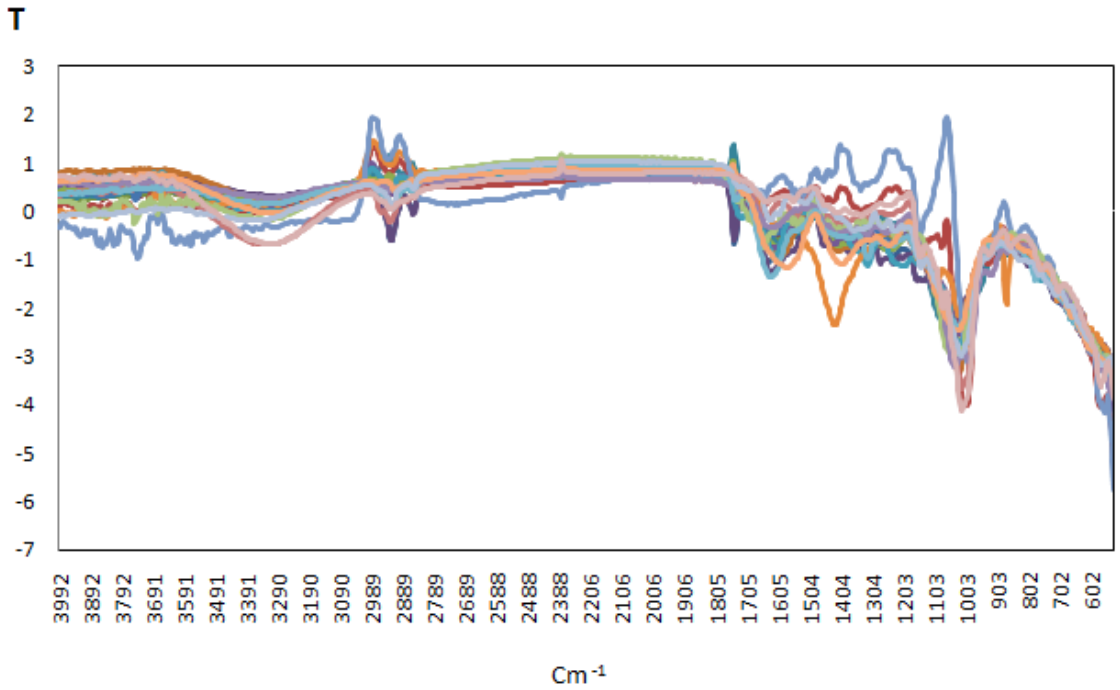

Figure 2. The average ATR-IR spectral data for 20 herbal samples (see Table 1) after performing standard normal variety preprocessing (n=3)

classification methods were able to discriminate the two medicinal seeds, *Hyocyamus niger* (no. 1) and *Peganum harmala* (no. 4) from other herbal samples (figrues 3-5). It is worth noting that the seeds used in this study were from different families of Solanaceae and

Zygophyllaceae. However the common feature of the seeds such as starch, oil, and protein composition could be the basis for their classification in one group. Similarly, the starch samples were clustered in a reasonable method. In HCA as displayed in figure 3, one cluster included three types of starch samples: *Zea mays* (no. 8), *Oryza sativa* (no. 14) and *Triticum aestivum* (no. 20). This cluster was divided into two subgroups with *T. aestivum* and *O. sativa* in one and *Z. mays* in the other. This pattern of classification was in accordance with the botanical studies for the similarity of *T. aestivum* and *O. sativa* in terms of their composition and origin. A similar pattern of classification was also seen in SOM method. On the contrary, K-means as displayed in table 2, was not able to discriminate between *Z. mays*, *O. sativa* and *T. aestivum* starches and all samples were clustered in one group. All the four classification methods were able to separate *Solanum tuberosum* (no. 13) starch from other starch samples. This pattern of scattering could be attributed to different origin of *S. tuberosum* starch as tuber from Solanaceae family *vs* the other starch samples that belong to Poaceae family. In case of leaves, it was observed that *Adiantum* (no. 12) a divergent sample of the leaves in this data set was well separated by the three methods K-means, HCA and PCA. In SOM (figure 5) this sample was co-clustered with two other leaf samples. The two leaf samples *Origanum vulgare* (no. 9) and *Melissa officinalis* (no. 10) belong to Lamiaceae family. Only two of the used classification methods in this study, HCA and SOM, were able to classify these two samples in one category. As seen in table 2, K-means clustering method was not successful with this issue. The two flowers in this study *Crocus sativus* (no. 15) and its adulterant *Carthamus tinctorius* (no. 7) were identified by the three methods PCA, HCA and SOM as different categories.

Table 2. Analysis of the 20 herbal samples based on K-means.

Class	Sample no.[1]
1	13
2	12
3	1,4,10,11,17,18
4	2,3,7,15,19
5	5,6,9,16
6	8,14,20

[1]For characterization of herbal numbers see table 1.

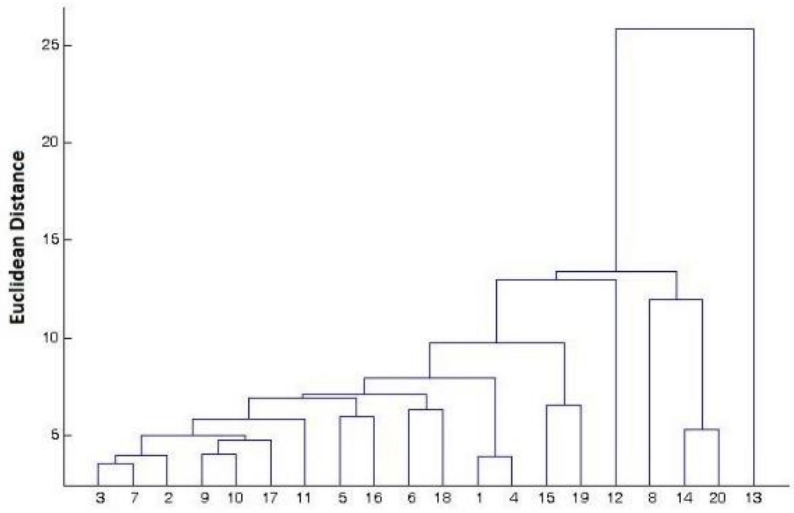

Figure 3. HCA analysis of the 20 herbal samples**.** For characterization of herbal numbers see table 1.

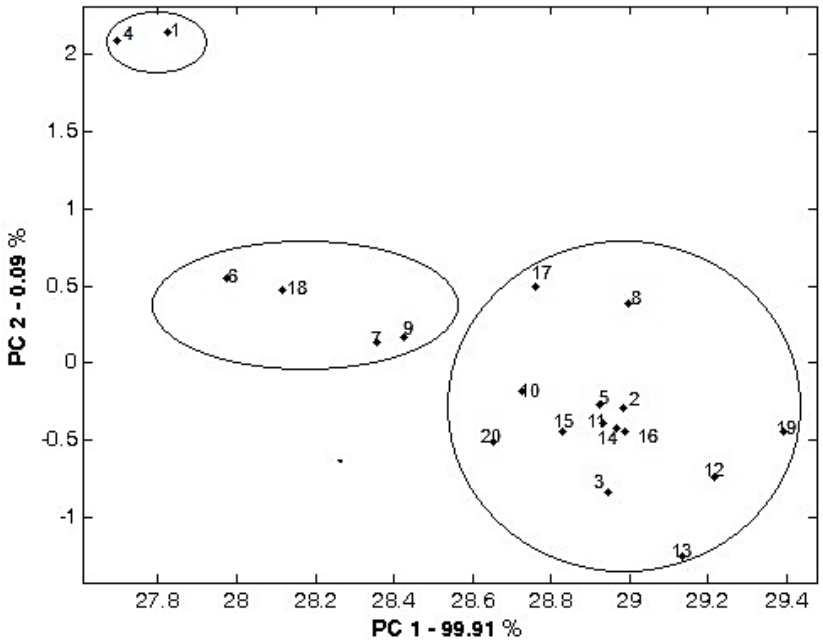

Figure 4. Cluser analysis of 20 herbal samples based on Principal Component Analysis (PCA). For characterization of herbal numbers see table 1.

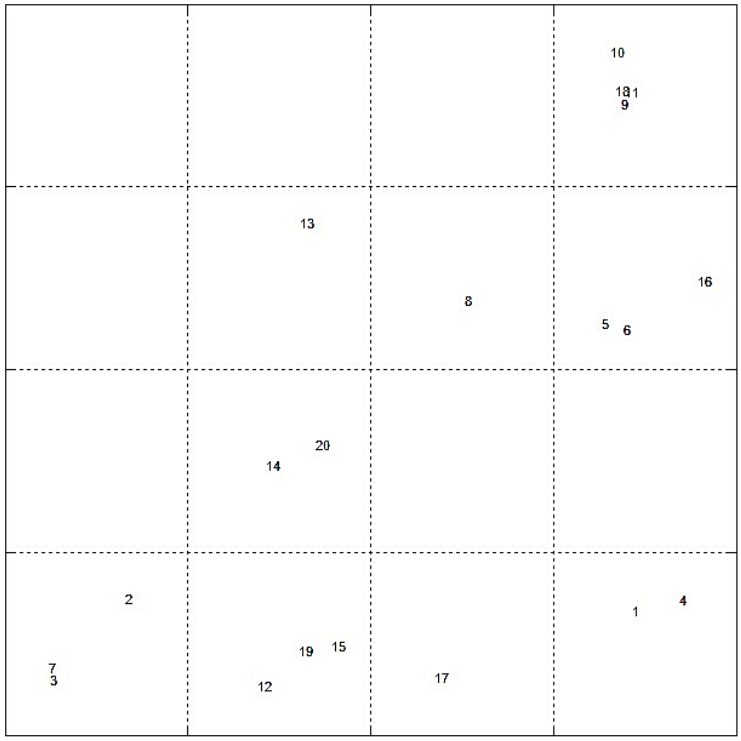

Figure 5. Classification of 20 herbal samples based on Kohonen self-organizing maps (SOM). For characterization of herbal numbers see table 1.

As described before this set of data is a random society of the medicinal plants which are normally used as reference in laboratory microscopic analysis. Microscopic analysis is a time consuming technique which needs special requirement and expertise. In addition the cross contamination of the samples by organelles from other samples adds to its complexity. For this purpose, different atlases of microscopic analysis such as American Herbal Pharmacopeia: botanical pharmacognosy and pulver atlas der drogen have been published for identification of the plant samples. Based on the results of this study it is possible to develop special libraries of ATR-IR for medicinal plants.

By exploring more reasonable models of classification, it would be possible to replace fingerprinting approaches such as IR as applicable and reproducible methods in identification of the plants. As a limitation, despite the reasonable described patterns, no other meaningful observation was observed in this study. These techniques could be largely used in quality control and detection of adulterations in herbal preparations. The results of this study are also suggestive of developing supervised classification models for identification of herbal samples. The subjective descriptions of the materials including macroscopic appearance, organoleptic characters, microscopic characteristics, and presence of the chemical substances could be replaced by the objective approach of ATR-IR. The consequence of this study is also in accordance with other studies reported [1,3] for the applicability of ATR-IR in herbal quality control.

Microscopical assays, microchemical tests orchromatography techniques are the most common techniques for identification and quality control of medicinal plants. Since such techniques are often expensive, nonspecific and time consuming, it is of great importance to find novel simple methods with more efficiency and specificity. Some studies on identification and quality control of medicinal plants' samples based on NMR data have been recently reported. In these methods spectral data are converted to fingerprints for each sample. The resulted matrix of data is thereafter subjected to further classification methods such as principal component analysis (PCA). The plot of the first two PCs (Principal Components) represents the plant samples in 2D space. The samples with more similarity are accordingly clustered within near distances. ATR-IR (Attenuated Total Reflectance Infra-Red) is a facile non-invasive method which is normally used in qualitative and quantitative analysis.

Acknowledgments

Financial support for this study was provided by the Pharmaceutical Sciences Research Center, Shiraz University of Medical Sciences.

Declaration of interest

The authors declare that there is no conflict of interest. The authors alone are responsible for the content of the paper.

References

[1] Cordell GA, Colvard MD. Natural products and traditional medicine: turning on a paradigm. *J Nat Prod.* 2012; 75(3): 514-525.

[2] Kokalj M, Kolar J, Trafela T, Kreft S. Finding the most appropriate IR technique for plant species identification. *Planta Med.* 2011; 77: 38.

[3] Turner NW, Cauchi M, Piletska EV, Preston C, Piletsky SA. Rapid qualitative and quantitative analysis of opiates in extract of poppy head via FTIR and chemometrics: towards in-field sensors. *Biosens Bioelectron.* 2009; 24(11): 3322-3328.

[4] Lai Y, Ni Y, Kokot S. Classification of raw and roasted Semen Cassiae samples with the use of Fourier transform infrared fingerprints and least squares support vector machines. *Appl Spectrosc.* 2010; 64(6): 649-656.

[5] Deconinck E, Cauwenbergh T, Bothy JL, Custers D, Courselle P, De Beer JO. Detection of sibutramine in adulterated dietary supplements using attenuated total

reflectance-infrared spectroscopy. *J Pharmaceut Biomed.* 2014; 100: 279-283.

[6] Samadi, Theodoridou K, Yu P. Detect the sensitivity and response of protein molecular structure of whole canola seed (yellow and brown) to different heat processing methods and relation to protein utilization and availability using ATR-FT/IR molecular spectroscopy with chemometrics. *Spectrochim Acta A.* 2013; 105: 304-313.

[7] Ghasemi Dehkordi N. *Most commonly used medicinal plants: microscopic and macroscopic quality control.* Isfahan: Chaharbagh Pub., 2014.

[8] Upton R, Graff A, Jolliffe G, Länger R, Williamson E. *American herbal pharmacopoeia: botanical pharmacognossy-microscopic characterization of botanical medicines.* Boca Raton: CRC Press, 2011.

[9] Eschrich W. *Pulver-atlas der drogen der deutschsprachigen arzneibücher.* Stuttgart: Deutscher Apotheker Verlag, 2003.

[10] Li T, Zhang H. Application of microscopy in authentication of traditional Tibetan medicinal plants of five *Rhodiola* (Crassulaceae) alpine species by comparative anatomy and micromorphology. *Microsc Res Techniq.* 2008; 71(6): 448-458.

[11] Steinley D. K-means clustering: a half-century synthesis. *Br J Math Stat Psychol.* 2006; 59(Pt 1): 1-34.

[12] Li HJ, Jiang Y, Li P. Characterizing distribution of steroidal alkaloids in *Fritillaria* spp. and related compound formulas by liquid chromatography-mass spectrometry combined with hierarchial cluster analysis. *J Chromatogr A.* 2009; 1216(11): 2142-2149.

[13] Sundaram J, Park B, Hinton A, Yoon SC, Windham WR, Lawrence KC. Classification and structural analysis of live and dead *Salmonella* cells using Fourier transform infrared spectroscopy and principal component analysis. *J Agric Food Chem.* 2012; 60(4): 991-1004.

[14] Milano Chemometrics and QSAR Research. Kohonen and CP-ANN toolbox for MATLAB, version 2.0.