ULTRAVIOLET-VISIBLE (UV-VIS) SPECTROSCOPY AND CLUSTER ANALYSIS AS A RAPID TOOL FOR CLASSIFICATION OF MEDICINAL PLANTS

SIMONA CODRUȚA AURORA COBZAC^{a,b,*}, DORINA CASONI^{a,b}, MIHAELA BADEA^c, BILJANA BALABANOVA^d, NATALIJA MARKOVA RUZDIK^e

ABSTRACT. The ultraviolet-visible (UV-Vis) spectroscopy coupled with cluster analysis (CA) was evaluated for the classification of some medicinal plants of different geographical growing area. To have a deeper view, the experiment was carried out on herbs belonging to different families. The UV-Vis spectra of hydroalcoholic extracts were acquired in the range of 200-800 nm. The hierarchical clustering analysis (HCA) was applied to the data matrix provided by unprocessed, normalized and standardized spectra respectively. Different types of distance measuring of (dis)similarity between the samples as well as different kinds of linkage or amalgamation rule were taken into account. The best results for the classification of the selected medicinal plants were obtained using Ward's method as the amalgamation rule combined with 1-Pearson r clustering distance measurement. The obtained results reveal the ability of HCA with Ward and 1-Pearson r algorithm to identify plant species even when the raw material has different provenience areas and different pedoclimatic growing conditions. In addition, this methodology revealed a direct link between herbs from different families.

Keywords: medicinal plants, classification/identification, UV-Vis spectroscopy, cluster analysis

^a Babeş-Bolyai University, Faculty of Chemistry and Chemical Engineering, 11 Arany Janos str., RO-400028, Cluj-Napoca, Romania

^b Research Center for Advanced Chemical Analysis, Instrumentation and Chemometrics – ANALYTICA, Babeş-Bolyai University 11 Arany Janos str., RO-400028, Cluj-Napoca, Romania

^c Faculty of Medicine, Transilvania University of Brasov, Nicolae Balcescu St., No. 56, Brasov, Romania

^d Department of Soil Chemistry and Hydrology, Faculty of Agriculture, Goce Delcev University, Krste Misirkov 10-A, 2000, Stip, Republic of North Macedonia

^e Department of Plant Production, Faculty of Agriculture, Goce Delcev University, Krste Misirkov 10-A, 2000, Stip, Republic of North Macedonia

^{*} Corresponding author: csimona@chem.ubbcluj.ro

INTRODUCTION

The World Health Organization (WHO) encourages and promotes the integration of traditional medical practices into healthcare systems, and increasingly supports the countries producing medicinal plants [1]. Traditional medicine is often termed complementary or alternative medicine (CAM) and according to the WHO estimation, about 80% of the emerging world's population relies on herbal remedies (drugs) for therapy [2, 3].

Usually, the term *herbal drugs* denote medicinal plants or plant parts converted into phytopharmaceutical products (teas, tablets, tinctures and extracts) by means of simple processes involving harvesting, drving, storage. mincing/grinding, compressing or extraction of active compounds with different solvents. The parts of medicinal plants that may be used are root, leaf, flowers, fruit, seeds or even the whole plant. Regarding the composition, phytopharmaceuticals are mixtures of many compounds with the specific physiological effect that may act individually, additively, or in synergy to improve health and cure different ailments. The phyto -compounds that are responsible for the biological characteristics of plant species are the secondary metabolites. It is well known that for the same species, differences in geographical location and growing conditions may alter the composition by increasing/decreasing the content of the constituents responsible for the phytotherapeutic effect or appearance of other compounds. Also, different parts of the same plant contain different amounts of secondary metabolites. In this context, methods for correct identification/authentication and classification of plants with specific therapeutic actions are acquiring higher importance for consumers, producers, and the pharmaceutical industry. Usually, both identification and classification processes depend on the identification of physical, chemical and biochemical differences and similarities between samples [4].

Numerous analytical techniques have been developed for quality control and evaluate compliance with the standard requirements. Instead of this, the methods for classification and identification/authentication were less approached, because of the need for advanced methods for comparison of the analytical results with a huge volume of information from a database. Exploratory data analysis, unsupervised and supervised techniques are often used as pattern recognition methods for the classification of herbal products and reveal the similarities and dissimilarities of sample properties. The aim of the unsupervised pattern recognition methods is to detect similarities whereas supervised techniques make use of calibration or training sets with a priori known the information to build a classification model. Cluster analysis (CA) and principal component analysis (PCA) are widely applied for authenticating or classifying plants based on spectral data [5, 6]. The PCA method allows visualization of the difference between samples, whereas CA can classify

objects based on their characteristics. The combination of CA and PCA has been widely used in current quality assessment, classification and identification/ authentication of herbs extracts is carried out based on a chromatographic fingerprint [10-14]. Spectral analysis including ultraviolet-visible (UV-Vis) and infrared (IR) spectroscopy can also be used for these purposes [7, 15-17]. Their advantages consist in avoiding the time consumed by chromatographic separation while retaining sufficient information to obtain a correct classification [18]. The methods are based on the fact that samples with similar spectral responses have the same chemical and pharmaceutical properties. Starting from the spectral results, the application of pattern recognition methods is the first step of the data analysis for sample classification. The cost-effectiveness of spectroscopy combined with chemometrics was confirmed by numerous literature studies [19, 20]. Also, spectral data analyzed by chemometric methods better exploit the relationship between the plant spectra and properties [8].

Therefore, the aim of the present work was to identify the most appropriate cluster analysis methodology able to provide a rapid identification/ classification of medicinal plants based on the UV-Vis spectra. For these investigations, seventeen medicinal plants belonging to various families and collected from Macedonia, Hungary and Romania were used.

RESULTS AND DISCUSSION

UV-Vis spectra of plant materials

UV-Vis spectral profile of hydroalcoholic extracts can provide qualitative and quantitative information upon the presence of different types of polyphenolic constituents [21, 22] that are the most common phytochemicals present in plants. The class of polyphenols includes phenolic acids (hydroxybenzoic acid derivates and hydroxycinnamic acid derivates), flavonoids (flavonols, flavones, flavanols, flavones, anthocyanidins, isoflavone), coumarins, stilbenes, tannins and lignans with specific absorbance values in the ultraviolet and visible spectral region.

A number of seventeen medicinal plants from spontaneous flora from Macedonia and different cultivars from Hungaria and Romania (Table 1) were selected for this study. The UV-Vis absorption spectra of their hydroalcoholic extracts were acquired in the range 200 – 800 nm. The obtained spectra profiles are presented in Figure 1. Two absorption bands were observed in the range from 200 to 250 nm and from 250 to 300 nm and a third broad absorption band is located between 300 and 400 nm. These absorption bands are related to different phenolic compounds present in the plant extracts that contain aromatic conjugated systems and absorb light in different UV spectral regions depending on the conjugations of the benzene rings.

No	Common name	Species	Part of the plant	Familly	Country of	Cada
					provenience	Coae
			-		(cultivar)	1
1	Liquorice	Glycyrrhiza glabra L.	radix*	Fabaceae	Macedonia	Lqi
			radix		Macedonia	
			radix**		Romania (Fares)	Lq3
2	Dandelion	officinale F.H.Wigg.	radix**	Asteraceae	Macedonia	Da4
			radix**		(Gyogyfu)	Da5
3	Comfrey	Symphytum officinale L.	radix**	Boraginaceae	Romania (Fares)	Co6
4	Valerian	Valeriana officinalis L.	radix**	Caprifoliaceae	Macedonia	Va7
			radix**		Romania (Fares)	Va8
5	Sweet fennel	Feniculum vulgare var. dulcis Mill.	fructus	Apiaceae	Macedonia	Fe9
			fructusTBC		Macedonia	Fe10
			fructus		Romania (Plafar)	Fe11
6	Dog rose	Rosa canina L.	fructus***	Rosaceae	Macedonia	Dr12
			fructus***		Macedonia	Dr13
			fructus**		Romania (Fares)	Dr14
7	Juniper	Juniperus communis L.	fructus***	Cupresaceae	Macedonia	Ju15
			fructus***		Romania (AdNatura)	Ju16
8	Bean	Phaseolus vulgaris L.	pericarpum	Fabaceae	Macedonia	Be17
			pericarpum		Romania (Fares)	Be18
9	Elderberry	Sambucus nigra L.	flos**	Adoxaceae	Macedonia	Eb19
			flos**		Romania (Fares)	Eb20
10	Wild thyme	Thymus serpyllum L.	herba**	Lamiaceae	Macedonia	Wt21
			herba**		Macedonia	Wt22
			herba**		Romania (Plafar)	Wt23
11	Horsetail	Equisetum arvense L.	herba***	Equisetaceae	Macedonia	Ht24
			herba**		Romania (Fares)	Ht25
12	Oregano	Origanum vulgare L.	herba**	Lamiaceae	Macedonia	Or26
			herba**		Romania (Fares)	Or27
13	Peppermint	Mentha x piperita L.	folium**	Lamiaceae	Macedonia	Pm28
			herba**		Romania (Fares)	Pm29
			folium**		Romania	Pm30
14	Blackberry	Rubus fructicosus L.	folium**	Rosaceae	Macedonia	Bb31
			folium**		Romania (Plafar)	Bb32
15	St.John's wort	Hypericum perforatum L.	herba***	Hypericaceae	Macedonia	Sw33
			herba**		Macedonia	Sw34
			herba**		Romania (Fares)	Sw35
16	Agrimony	Agrimonia eupatoria L.	herba**	Rosaceae	Macedonia	Ag36
			herba**		Romania (Fares)	Ag37
17	Stinging nettle	Urtica dioica L.	semen	Urticaceae	Macedonia	SnS38
			folium**		Macedonia	SnF39
			herba**		Romania (Fares)	SnH40

Table 1. The name and provenience of the medicinal plants

*-pulvis(powder); **-concis(pieces); ***-totus(entire)

Among the polyphenolic constituents usually present in medicinal plant extracts, different flavonoid subclasses show different shapes of spectra. According to the obtained spectral profile, most of the selected extracts are rich in flavonoids characterized by absorption band at around 240–290 nm (known as Band II) and flavones and flavonols characterized by absorption maximum around 310–370 nm (generally a longer wavelength for flavonols than for flavones).



Figure 1. The absorption UV-Vis spectra of the investigated medicinal plant.

The absorption spectra profiles acquired in the range of 200-800 nm revealed that specific differences and similarities between the investigated samples are difficult to distinguish. Therefore, the cluster analysis was conducted to investigate the (dis)similarities in a spectral dataset and classify samples based on their characteristics.

S. C. A. COBZAC, D. CASONI, M. BADEA, B. BALABANOVA, N. MARKOVA RUZDIK

Cluster analysis

Among the general categories of cluster analysis methods, hierarchical cluster analysis (HCA) with joining or tree clustering algorithm was applied on the digitized data matrix (40 samples x 601 variables) from unprocessed. normalized and standardized spectra respectively. The Euclidean. Squared Euclidean, City-block (Manhattan), Chebychev, Power, Percent disagreement and 1-Pearson r distance HCA algorithms were used for clustering distance measurements. Also, different linkage or amalgamation rule including Single linkage (nearest neighbor), Complete linkage (furthest neighbor), Unweighted pair-group average (UWPGA), Weighted pair-group average (WPGA), Unweighted pair-group centroid (UWPGC), Weighted pair-group centroid (WPGC). and Ward's method were applied in order to decide when two clusters are sufficiently similar to be linked together. The commonly applied, Single (SL) and Complete linkage (CL) algorithms used for clustering classification have not provided satisfactory results (Figure 2 a,b). In our case, the best results were obtained using Ward's amalgamation algorithm. when the highest similarity percent were revealed by short linkage distances (Figure 2, c).



Figure 2. Dendrogram of investigated medicinal plant based on the UV-Vis spectral data using different linkage method (a-SL; b-CL; c-Ward's method) and Euclidian distance for clustering

In terms of distance measurement, among *Euclidean Square*, *Euclidean*, *City-block*, *Chebyshev*, *Power SUM*($ABS(x-y)^2$)^{1/r} and 1-Pearson r applied methods, the 1-Pearson r method was the most effective (Figure 3) since herbs from the same species were grouped together.



Figure 3. Dendrogram for classification of investigated medicinal plants based on the UV-Vis spectral data using *Ward's* method as linkage algorithm and different methods for distance measuring: (**a**)-*City-block (Manhattan)* distance; (**b**)-*Chebyshev*; (**c**)-1-*Pearson r*

The same protocol of statistical evaluation was applied to data obtained from spectra after their normalization and standardization. In comparison with the results obtained from unprocessed UV-Vis spectra, a better clustering of plants from the same species could be observed. Nevertheless, different clustering was highlighted depending on the method used for spectra processing (Figure 4).

It is noteworthy that using the combination Ward's / 1-Pearson r, the same clustering of the samples was achieved both for unprocessed and processed spectra (Figure 5). The *Ward's* method is distinct from all other methods because it uses an analysis of variance approach to evaluate the distances between clusters and attempts to minimize the sum of squares of

any two (hypothetical) clusters that can be formed at each step. Moreover, the use of *1-Pearson r* distance is more efficient to find the similarity, providing small sub-clusters especially in case of correlated data such as those obtained from the UV-Vis spectra of the medicinal plant's extracts. So, we can conclude that if the input matrix is a correlation matrix, data would be transformed as *1-Pearson r* distances before the analysis.



Figure 4. Classification dendrogram of investigated medicinal plants based on normalized (\mathbf{a} ; \mathbf{b}) and standardized (\mathbf{c} , \mathbf{d}) spectra using *WPGA/Euclidian* (\mathbf{a} , \mathbf{c}) and *Ward's/City-block* (\mathbf{b} , \mathbf{d}) used as linkage algorithm/distance measuring mode

According to the obtained dendrogram (hierarchical tree clustering) (Figure 5), the investigated samples are divided into three principal clusters. A clear classification of the investigated plants according to the family could not be observed, but similarities in the composition of the plants belonging to different families could be taken into account. The most important result highlights a very good classification of the samples according to the species, despite their different geographical provenience area. Twelve of the investigated plant species (liquorice-Lq, dandelion-Da, dog rose-Dr, juniper-

Ju, bean-Be, elderberry-Eb, wild thyme-Wt, horsetail-Ht, St.John's wort-Sw, agrimony-Ag and stinging nettle-Sn) were classified in the individual subclusters. Plants rich in volatile oils, such as fennel-Fe, valerian-Va and peppermint-Pm were not grouped into sub-clusters but are included in the same principal cluster. Differences in spectra profiles could be assigned to possible variation in volatile oil content. An interesting clustering was observed in the case of blueberry-Bb and oregano-Or that were grouped according to the geographical growing area. This can be explained based on some similarities in their spectra.



Figure 5. The dendrogram (hierarchical tree clustering) of the relationship between the investigated medicinal plant extracts based on the UV-Vis spectroscopy data using the *1-Pearson r* distance as a measure of (dis)similarity between the samples and *Ward's* method as linkage or amalgamation rule specifying the (dis)similarity between the clusters

Although the HCA method has been widely used in the common approach for the medicinal plant's evaluation [23], to our best knowledge there is not concluded which would be the most appropriate combination of distance measurements/linkage methodology that can be used HCA analysis in order to provide a rapid identification / classification of the plant species even in cases in which the plant comes from different places and different

S. C. A. COBZAC, D. CASONI, M. BADEA, B. BALABANOVA, N. MARKOVA RUZDIK

pedoclimatic conditions. During the investigations developed in the present study, it was demonstrated for the first time that using the *Ward's / 1-Pearson r* combination in HCA analysis of UV-Vis spectra, a rapid classification of plants can be obtained according to the species. Moreover this methodology has the ability to generally overcomes the influences due to the pedoclimatic conditions, those related to the sample preparation (such as the degree of grinding) or those related to the use of the different parts of the plant. Examples can be revealed by classification of *Dog rose* (Dr) samples which are of different grinding and *Stinging nettle* samples that are of different part of plant provenience.

CONCLUSIONS

The ability of UV-Vis spectroscopy with hierarchic cluster analyses (HCA) was investigated for the rapid classification of medicinal plants. Using *Ward's method* as amalgamation rule combined with *1-Pearson r* for distance measurement, forty samples of medicinal plants from various families and of different geographical area provenience were grouped into individual subclusters according to their belonging species. The applied methodology revealed the same classification for the investigated samples in all cases using the original, normalized and standardized data provided by UV-Vis spectra respectively. Based on the obtained results, the proposed pattern recognition methodology combined with UV-Vis spectroscopy can be used for the rapid identification of plant species from different geographical areas.

EXPERIMENTAL SECTION

Materials and methods

The study was carried out on seventeen medicinal plants belonging to different families and having different European provenience areas (Macedonia, Hungary, and Romania). The Hungarian and Romanian vegetal material was purchased from a specialized store as phytoproducts for infusion. All phytoproducts are certified regarding the species and purity in concordance with regulations of Romanian and Hungarian Pharmacopoeias. The plants originated from North Macedonia were collected from three localities in the Osogovo mountains basin situated in the south-eastern part of the country. They were identified by determination key using the data from Matevski [24], a specimen is kept in the herbarium at the Department of Plant Production,

Faculty of Agriculture, Goce Delcev University in Shtip, Republic of North Macedonia. Depending on the part of plants that have a recognized therapeutic effect, forty hydroalcoholic extracts were obtained from roots (*radix*), leaves (*folium*), flowers (*flos*), fruits (*fructus*), seeds (*seeds*) or entire aerial parts (*herba*) (Table 1). The raw material was powdered using a mill, except sample Dr13 which was ground to coarse dimension. For each sample, 2 g of minced material was extracted with 20 mL of ethanol - water (7:3, v/v) mixture by a ten days maceration at room temperature. The extract was separated by centrifugation and the remained solid residue was two times centrifuged with 2 mL ethanol - water (7:3, v/v). The combined extracts were adjusted to 25 mL volumetric flask using a mixture of ethanol - water (7:3, v/v) in each case.

UV-Vis spectra acquisition and preprocessing

The molecular absorption UV-Vis spectra were recorded using the Jasco V-550 double beam spectrophotometer (Jasco Corporation, Japan). Spectra of the previously resulted extracts were collected in the region of 200 - 800 nm with 1 nm acquisition interval after 1:100 dilution. To remove any irrelevant information several spectral preprocessing techniques were applied before spectral data analysis. Usually scatter or baseline correction, signal enhancement and statistical filtering of signal noise preprocessing methods can be applied to the original spectra [25]. In this case, smoothing algorithms including the Savitzky-Golay method was used as a low-pass filter to correct additive and multiplicative effects in the spectral data and preserve the important features of information by removing random variations in spectra. Secondly, normalization and standardization processing techniques were applied. The standardization procedure was based on the transformation of initial spectral data by subtracting the mean from each value and dividing it by the standard deviation. All spectral values were replaced by standard values computed as Standardized value = (raw value - mean) / Standard deviation. Normalization was applied to eliminate the shifting of the peaks on the wavelength scale. For this the formula $(A-A_{min})/(A_{max}-A_{min})$, where A is absorbance value, Amin is minimum value and Amax is the maximum value of absorbance in the spectra.

Spectral data analysis

Cluster analysis (CA), an unsupervised pattern recognition method was used in order to detect similarities in the spectral dataset and classify the selected medicinal plants based on quantitative characteristics. The clustering techniques are divided into hierarchical and nonhierarchical type.

In the common approach for the medicinal plant's evaluation, the most popular clustering technique is hierarchical clustering analysis (HCA) due to the large flexibility to alter the similarity measurement criteria and the applied linkage methods to suit different applications. HCA technique is based on the creation of branched structures, called dendrograms which are qualitative in nature and permit visualization of clusters and correlations amongst samples. In order to decide when clusters should be merged or split, a measure of (dis)similarity between the samples as well as a linkage criterion specifying the (dis)similarity between the clusters is used. The result of this type of clustering is the hierarchical tree where the horizontal axis denotes the linkage distance between the obtained individual clusters. In this study, the Euclidean, Squared Euclidean, City-block (Manhattan), Chebychev, Power and 1-Pearson r distance were used for distance measurements. The Single Linkage, Complete Linkage, Unweighted Pair-Group Average, Weighted Pair-Group Average, Unweighted Pair-Group Centroid, Weighted Pair-Group Centroid (median). and Ward's method were used as different linkage or amalgamation rule in order to determine when two clusters are sufficiently similar to be linked together.

REFERENCES

- 1 WHO; WHO Guidelines on Good Agricultural and Collection Practices (GACP), 2003
- 2 WHO; Traditional Medicine. Growing Needs and Potential. WHO Policy Perspective on Medicine. WHO, Geneva, 2002
- 3 A. Gurib-Fakim; Mol. Aspects Med., 2006, 27, 1-93
- 4 O.Y. Rodionova; A.V. Titova; A.L. Polmerantsev; *Trends Anal. Chem.*, **2016**, *78*, 17-22
- 5 M. Forina; P. Oliveri; S. Lanteri; M. Casale; *Chemom. Intell. Lab. Syst.*, **2008**, 93, 132-148
- 6 Y. Wei; W. Fan; X. Zhao; W. Wu; H. Lu; Anal. Lett., 2015, 48, 817-829
- 7 Z. Cao; Z. Wang; Z. Shang; J. Zhao; PLOS ONE, 2017, 1-14
- 8 H.A. Gad; S.H. El-Ahmady; M.I. Abou-Shoer; M.M. Al-Azizi; *Phytochem. Anal.*, **2013**, *24*, 1-24
- 9 I.M. Simion; C. Sârbu; *Spectrochim. Acta A: Mol.Biomol.Spectroscopy*, **2019**, Accepted manuscript, http://doi.org.10.1016/j.saa.2019.04.038
- 10 J. Fibigr; D. Satinsky; P. Solich; Anal. Chim. Acta, 2018, 1036, 1-15
- 11 S.S. Chavan; V.M. Jadhav; V.J. Kadam; *Int. J. Pharm. Sci. Rev. Res.*, **2017**, *43*, 161-168

- 12 X. Zhang; J. Zhang; S. Zhang; Z. Qian; S. Wu; J. Liu; Y. Ye; J. Si; *Ind. Crops Prod.*, **2018**, *124*, 707-718
- 13 G. Alvarez-Rivera; D. Ballesteros-Vivas; F. Parada-Alfonso; E. Ibañez; A. Cifuentes; *Trend. Anal. Chem.* **2019**, *112*, 87-101
- 14 M. Simion; D. Casoni; C. Sarbu; J. Pharm. Biomed. Anal. 2019, 163, 137-143
- 15 C. Cheng; J. Liu, W; Cao, R. Zheng; H, Wang; C. Zhang; *Vib. Spectrosc.*, **2010**, *54*, 50-55
- 16 T. Hu; W.Y. Jin; C.G. Cheng; Spectroscopy, 2011, 25, 271-285
- 17 A. Dankowska; W. Kowalewski; Spectrochim. Acta A, Mol. Biomol. Spectroscopy, **2019**, *211*, 195-202
- 18 D.D. Joshi; Herbal Drugs and Fingerprints. In *UV-Vis Spectroscopy: Herbal Drugs and Fingerprints*, Springer, New Delhi Heidelberg, 2012, Chapter 6, pp 101-105
- 19 L.A. Berrueta; R.M. Alonso-Salces; K. Heberger; J. Chromatogr. A, 2007, 1158, 196-214
- 20 J. Moros; Trends Anal. Chem. 2010, 29, 578-591
- 21 S.M. Dhivya; K. Kalaichelvi; Int. J. Curr. Pharm. Res., 2017, 9, 46-49
- 22 K. Kalaichelvi; S. M. Dhivya; Int. J. Herb. Med., 2017, 5, 40-44
- 23 D. Jing; W. Deguang; H. Linfang; C. Shilin; Q. Minjian; *J. Med. Plants Res.*, **2011**, *5*, 4001-4008
- 24 V. Matevski; Flora of the Republic of Macedonia, 1st Edition, Macedonian Academy of Science and Arts, Skopje, 2010 (in Macedonian)
- 25 A. Rinnan; F.W.J. van den Berg; S.B. Engelsen; *Trends Anal. Chem.*, **2009**, *28*, 1201-1222