Chemometrics and Intelligent Laboratory Systems 235 (2023) 104783

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Authentication of soothing herbs by UV–vis spectroscopic and chromatographic data fusion strategy

Clara Pérez-Ràfols [a,b,*], Núria Serrano [a,b,**], José Manuel Díaz-Cruz [a,b]

[a] Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona (UB), Martí i Franquès 1-11, 08028, Barcelona, Spain
[b] Water Research Institute (IdRA), University of Barcelona (UB), Martí i Franquès 1-11, 08028, Barcelona, Spain

ARTICLE INFO

Keywords:
Soothing herbs
Data fusion
Chemometrics
Chromatography
Spectroscopy

ABSTRACT

A data fusion approach combining chromatographic and spectroscopic profiles is proposed for the discrimination and classification of soothing herbs in different types of herbal preparations. Particularly, chamomile, lavender, passionflower, and valerian were considered. The proposed data fusion approach revealed a higher clusterization ability than each analytical technique in a separate way, which was assessed through an exploratory analysis based on Principal Component Analysis (PCA) coupled to Silhouette analysis: percentage of samples with a negative Silhouette width were 19, 15 and 10 for chromatography, spectroscopy and data fusion, respectively. Furthermore, a Partial Least Squares – Discriminant Analysis (PLS-DA) model developed based on data fusion was able to perfectly discriminate samples of chamomile, passionflower, and valerian in a set of 20 samples, overcoming the difficulties related to dealing with different types of herbal preparations including pure herbs, infusions, tablets, capsules and herbal drops.

1. Introduction

Herbal medicines are herbs, herbal preparations or herbal derivative products that comprise as active ingredients plant parts, such as flowers, roots or leaves. They can be found as fresh or dried plants, teas, tablets, capsules, powders, and extracts, and are mainly used to treat or prevent mild to moderate illnesses [1]. Herbal medicines have their origins in ancient civilizations cultures, constituting an important part of the traditional medicine. Modern science has recognized their active action, and it has included in contemporary pharmacotherapy a variety of drugs of plant origin, identified by ancient cultures and used throughout the times [2,3]. Thus, although it is true that herbal medicines have been already used for centuries, nowadays there exists a growing demand for medicinal plants worldwide [4]. For instance, the market demand in China for herbal medicine has considerably increased, being the annual consumption of herbal medicine of more than 400,000 tons [5]. A more global analysis of the herbal medicine market suggests that it is expected to grow by $ 39.52 billion during 2022–2026, progressing at a compound annual growth rate (CAGR) of 6.69% over the analysis period [6].

Although herbal medicines are used as natural remedies and not strictly considered as drugs, it should be borne in mind that this does not guarantee that they are always safe and beneficial. Thus, quality control is of utmost importance and herbal medicines should undergo the testing to receive approval from the administration [7]. However, the existence of multiple preparations for the same type of herb (e.g., tablet, infusion, drop) increases the difficulty in their quality control, thus facilitating the possibility of frauds and tampering. This fact, together with the constant increase in the use of herbal medicines around the world, makes the characterization, identification and authentication of herbal medicines a subject of major concern both from a safety and quality control point of view [1]. Thus, the aim of this work is to develop a method able to discriminate among soothing herbs regardless of their preparation method. In particular, chamomile, passionflower, valerian root and lavender were considered in this study as the most common soothing herbs used to treat or prevent stress and anxiety, which are two recognized emotional disorders nowadays. In particular, chamomile infusions are commonly used as mild sedatives to reduce nerves and calm anxiety, and to treat insomnia, nightmares, hysteria, and other sleep complications [8]. Both passionflower and valerian roots are known to increase gamma-aminobutyric acid levels in the brain, promoting relaxation,

* Corresponding author. Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona (UB), Martí i Franquès 1-11, 08028, Barcelona, Spain.
** Corresponding author. Department of Chemical Engineering and Analytical Chemistry, Universitat de Barcelona (UB), Martí i Franquès 1-11, 08028, Barcelona, Spain.
E-mail addresses: claraperezrafols@ub.edu (C. Pérez-Ràfols), nuria.serrano@ub.edu (N. Serrano).

https://doi.org/10.1016/j.chemolab.2023.104783
Received 28 November 2022; Received in revised form 9 February 2023; Accepted 12 February 2023
Available online 13 February 2023

easing generalized anxiety, alleviating restlessness and aiding sleep [9,10]. Lavender has been revealed as beneficial to sleep, anxiety and overall mood [11].

Reported applications in the literature for the characterization of soothing herbs have usually been based on the use of chromatographic fingerprints. This approach is based on the chromatographic separation and identification of marker compounds from other components, and it is considered one of the most important and accepted approaches for quality assessment of herbal preparations [12–15]. Indeed, both World Health Organization (WHO) and Food and Drug Administration (FDA) have recognized chromatographic fingerprinting as a strategy to evaluate the quality of herbs and herbal preparations [16,17]. Nevertheless, fingerprints can also be acquired from spectroscopic techniques such as ultraviolet–visible (UV–Vis), infrared (IR) and Raman, relying on the fact that samples with comparable spectral responses have equivalent chemical and pharmaceutical properties [12,18,19]. To the best of our knowledge, the combination of chromatographic and spectroscopic techniques by data fusion to address this concern has not been described yet. Data fusion is an approach where the data from multiple sources of different nature are combined and analysed jointly in order to take advantage of their features and improve the representation of information compared to the respective sources separately [20]. Chromatographic and optical data can be fused using either a low-level or mid-level data fusion approach. In the former, each data set is first pre-processed independently to reduce noise and redundant information, and then the matrices are concatenated to obtain the augmented matrix. Nevertheless, this approach requires that both data sets have comparable magnitudes to avoid the domination of the larger data set in the subsequent analysis. Alternatively, a mid-level data fusion combines relevant features extracted from each data set independently, being Principal Component Analysis (PCA) the most popular method employed in this extraction [21].

The current work aims to compare the features of individual UV–Vis spectroscopic and chromatographic measurements to those achieved by the combination of both to go deeper in the characterization, identification and authentication of herbal medicines. Thus, in this study, spectroscopic and chromatographic data have been firstly analysed individually with pattern recognition methods in order to evaluate the capabilities of each analytical technique for the characterization, identification and authentication of chamomile, passionflower, valerian root and lavender in different commercial preparations. Both techniques provided complementary information that allowed a data fusion approach leading to an improvement of the classification model, thus demonstrating the benefits of data fusion in the characterization, identification and authentication of herbal medicines.

2. Experimental section

2.1. Samples and sample treatment

A total of 20 samples of herbs and herbal preparations of chamomile (5 samples: 1 herb, 4 infusion bags), passionflower (5 samples: 1 herb, 1 herbal drops, 2 tablets and 1 capsule), valerian (8 samples: 1 herb, 1 infusion bag, 1 herbal drops, 3 tablets and 2 capsules) and lavender (2 samples: 1 herb and 1 herbal drops) were purchased in herbalist's shops and big stores located in Barcelona (Spain). All considered samples are 100% pure and it should be noted that not all herbs could be found in all types of herbal preparations. Chamomile, passionflower, valerian and lavender were selected as four representative examples of soothing herbs to alleviate stress and anxiety.

Sample treatment was then performed as follows: 1.5 g of herb samples or one dose of herbal preparation sample were boiled in natural mineral water for 5 min, left to cool at room temperature, filtered through 0.2 μm nylon filters and stored at −4 °C until further analysis.

All samples were analysed in triplicate to account for the possible variability of the experimental procedure.

2.2. UV–vis spectroscopic measurements

UV–vis spectroscopic measurements were performed using a SPELEC equipment by Dropsens-Metrohm (Oviedo, Spain) and controlled by the software DropView 8400 from Dropsens-Metrohm. UV–vis measurements were carried out using two optical fibres (TFIBER-VIS-UV) with their respective collimators coupled to a sample holder CUV-UV from Ocean Insight (Orlando, USA). Samples were disposed in plastic cuvettes supplied by Eppendorf (Hamburg, Germany) and a cardboard cover was employed to protect the cuvette from stray light. Absorbance spectra were recorded with wavelengths between 200 and 900 nm in comparison with a blank of pure water.

Each herbal medicine sample was analysed in triplicate, resulting in a total of 60 UV–vis spectra.

2.3. Chromatographic measurements

HPLC-UV chromatograms were obtained using an Agilent 1200 Series instrument (Palo Alto, CA, USA) equipped with a quaternary pump (G1311A), a vacuum degasser (G1322A), an autosampler (G1329A) and an ultraviolet–visible detector (G1314B), and controlled with the Agilent ChemStation software package. Chromatographic fingerprints of considered soothing herbs extracts were acquired with a reverse phase Kinetex® C18 column (5 μm C18 100 Å, 100 × 4.6 mm) from Phenomenex (Torrance, CA, USA) and gradient elution using 0.1% formic acid in Milli-Q water (solvent A) and methanol (solvent B) as mobile phase components. The elution gradient program was as follows: 0–2 min, isocratic elution at 5% solvent B; 2–4 min linear gradient from 5 to 25% solvent B; 4–12 min, at 25% solvent B; 12–14 min, from 25 to 45% solvent B; 14–16 min, at 45% solvent B; 16–18 min, from 45 to 95% solvent B; 18–20 min, at 95% solvent B; 20–21 min, back to initial conditions at 5% solvent B; and 9 min keeping this composition of the mobile phase for column reequilibration [22]. The chromatographic column was kept at room temperature, and a mobile phase flow rate of 1 mL min^{-1} and an injection volume of 20 μL were employed. HPLC-UV chromatographic fingerprints were registered at 280 nm.

Each herbal medicine sample was injected in triplicate, except for the four pure herbs, which were injected in quadruplicate and along the sequence to be used as references for peak alignment. This generated a total of 64 chromatograms.

2.4. Data treatment

Both chromatographic and UV–vis spectroscopic data were retrieved from the instrument using the corresponding software and treated in a Matlab® environment [23]. PCA and Partial Least Squares – Discriminant Analysis (PLS-DA) models were constructed using PLS_Toolbox from Eigenvector Research [24], which is compatible with Matlab.

Chromatographic data were pre-treated prior to the application of chemometrics in order to avoid artefacts resulting from baseline drifts or small time shifts. Baseline was removed using the assisted baseline estimation and denoising using sparsity (BEADS) algorithm [25,26] with the following parameters: filter order = 1, cut-off frequency = 0.004 cycles/sample, asymmetry ratio = 17, amplitude = 0.1, $\lambda_0 = 0.1$, $\lambda_1 = 1$, and $\lambda_2 = 10$. Additionally, chromatographic peaks were aligned using the correlation optimized warping (COW) algorithm included in PLS_Toolbox.

As stated in section 2.2, UV–vis data were acquired with SPELEC in absorbance mode, using a blank of pure water and within the wavelength range 200–900 nm provided by the diode array of the instrument. This produced 2007 absorbance values per spectrum (*ca.* a value every 0.35 nm). However, a preliminary analysis of the spectra showed that the region most sensitive to the different types of herbs was between 250 and 500 nm. Then, only this part of the spectrum, including 700 measurements, was considered for data analysis.

PCA models built from a single analytical technique were con-

structed using either pre-treated chromatograms or UV–vis spectroscopic profiles, which were column-wise mean centred and scaled to unit standard deviation. The employed data sets had a size of 3021 time points x 64 samples and 700 wavelengths x 60 samples for chromatographic and spectroscopic data, respectively. For data fusion, the scores yielded by each sample in the two previous models for principal components 2 to 4 (HPLC) and 1 to 3 (UV–vis) were combined in a single matrix (6 PCA scores x 60 samples) that was further employed for the building of PCA and PLS-DA models. This can be understood as a mid-level data fusion strategy, since extracted features of the data (the scores) are used instead of the full preprocessed data [21]. A single matrix containing all samples was employed for PCA models. Clusterization in PCA models was analysed through Silhouette analysis [27], which was carried out by a homemade Matlab program. In particular, Silhouette width for a sample i was calculated using Equation 1, where $\overline{d(i,g)}$ is the average Mahalanobis distance from sample i to all samples belonging to the same cluster and $\overline{d(i,k)}$ is the average Mahalanobis distance from sample i to all samples belonging to other classes.

$$s_i = \frac{\overline{d(i,k)} - \overline{d(i,g)}}{\max(\overline{d(i,k)} - \overline{d(i,g)})} \qquad \textit{Equation } 1$$

For PLS-DA, samples were distributed into training (60%) and test (40%) sets, ensuring that all classes were evenly represented in both sets, except for lavender that was only included in the test set because the number of available samples was not high enough to reliably train the model. A PLS2 model based on binary 0/1 indexes was built considering chamomile, passionflower and valerian as classes. Class index matrix for the training set is reported in Table S1 and 3, latent variables (LV) were selected, corresponding to the first minimum in the average error classification from cross-validation (Fig. S1). For this purpose, cross-validation was based on venetian blinds, using 10 splits and 1 sample per split.

PLS-DA model was assessed by means of several classification parameters. Root mean square error (RMSE) is the standard deviation of the residuals, sensitivity reports the ratio of samples belonging to a considered class correctly classified as such, and specificity indicates the ratio of samples not belonging to a considered class correctly classified as "out of class". Additionally, non-error classification rate (%NER) is calculated for each class as an average between sensitivity and specificity, classification rate (%) is the percentage of true positive and negative samples with respect to the total samples and accuracy is calculated as a global parameter of the model that indicates the percentage of total samples correctly classified.

3. Results and discussion

Chromatographic fingerprints, as the most common and accepted approach for quality assessment of herbal preparations, were first acquired for different formulations of chamomile, passionflower, valerian and lavender. For this purpose, an HPLC method previously reported for the determination of polyphenolic compounds was employed [22] because soothing herbs are known to be rich in polyphenols [28]. Given the complex nature of chromatograms, data were first pre-treated to correct artefacts that could later distort the chemometric model. In this sense, baseline correction and peak alignment are crucial steps. Fig. 1a shows pre-treated chromatograms of four samples corresponding to the pure herbs of chamomile, lavender, passionflower, and valerian. Visual inspection reveals a large number of chromatographic peaks distributed along the time axis and with apparent differences among the four considered herbs, which indicates that the selected HPLC method may be, in principle, suitable for herbal identification. Nevertheless, a closer inspection of each class of herbal medicine revealed that different types of preparations (e.g. infusion, herbal drop, root, tablet or capsule) also result in significant changes in the chromatographic profile. As a representative example, Fig. 1b displays the pre-treated chromatograms acquired for valerian in different preparations. As it can be observed, capsule, tablet and infusion present some common peaks with clear differences in the concentration of the extracted substances. Moreover, other preparations (especially these extracted from the roots and herbal drops) show an obvious different distribution of peaks. Therefore, the

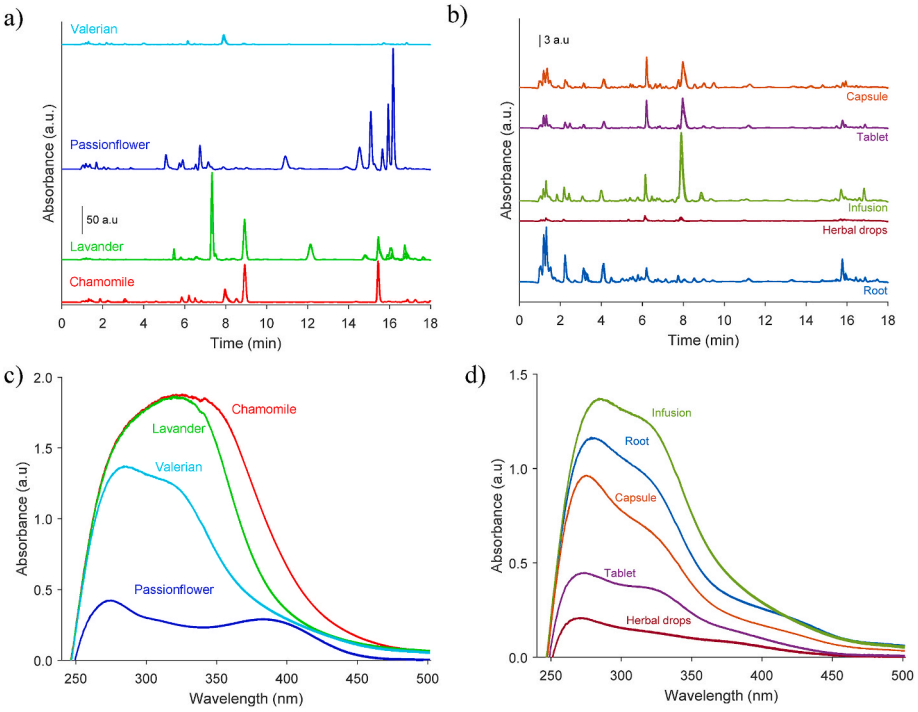

Fig. 1. Representative pre-treated chromatograms (a, b) and UV–vis spectra (c, d) of three replicates of chamomile (red), lavender (green), passionflower (blue) and valerian (cyan) pure herbs (a, c); and several preparations types of valerian (b, d). In chromatograms, baseline was adjusted with BEADS and peaks were aligned with COW. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

great variety in formulations available in the market represents a challenge in the correct characterization, identification and authentication of herbal medicines, which is expected as organic compounds will not always be extracted with the same efficiency.

Visual inspection was also carried out with spectroscopic fingerprints, which have also been reported for the authentication of medicinal plants [12,18,19]. Given the distinctive colours of the infusions obtained, UV–vis spectroscopic was selected for this purpose. Similar to what was observed in HPLC, significant differences could be visually detected in spectroscopic profiles of different herbs (Fig. 1c) but also among several types of preparations belonging to the same herbal medicine (Fig. 1d). Thus, it could be concluded from visual inspection that both chromatographic and spectroscopic profiles may be suitable for the characterization, identification and authentication of chamomile, passionflower, valerian root, and lavender. Therefore, both sets of data were considered for further evaluation by means of chemometric analysis.

Preliminary exploratory analysis was carried out using PCA on the full set of samples (20 samples with the corresponding replicates) with either chromatographic or UV–vis spectroscopic profiles. Fig. 2 shows both the diagram of scores and the Silhouette analysis plot obtained for each data set. Regarding the former, it can be observed that both techniques provide some clusterization among the 4 types of soothing herbs considered, with chamomile and valerian being the better separated group using HPLC (Fig. 2a) and spectroscopic (Fig. 2c) data, respectively. It should be noted that principal components (PC) 1 to 3 were all relevant to herbal differentiation when the optical data set was considered whereas PC1 in the model built from chromatographic data did not contribute to such differentiation, which was much clearer with PCs 2 to 4. In order to better assess the clusterization achieved with each model, Silhouette analysis was carried out. The Silhouette width is a quality measure that specifies how much closer each sample is to the cluster it has been assigned to in comparison to other clusters. The value of the Silhouette width may range from −1 to 1, with negative values indicating that the sample is actually closer to another cluster than to the one it was assigned to Ref. [27]. The results displayed in Fig. 2b and c shows that, no matter the data set employed, most samples presented a positive Silhouette width. More specifically, 81 and 85% of the samples for HPLC and UV–vis data, respectively, presented a positive Silhouette width (Table 1). This fact indeed agrees with the clusterization observed in the scores diagram. Regarding the measurements that yielded a negative Silhouette width, they mostly belong to all three replicates of one sample from each class (4 replicates in the case of chamomile using HPLC data as this sample corresponds to the herb), which manifests that the methodology is reproducible and that differences may be attributed to the sample. Only in a limited number of samples different Silhouette width signals among the replicates were observed but, in these cases, its absolute value was always low. Interestingly, comparing the results obtained using chromatographic and UV–vis spectroscopic data, the samples yielding negative Silhouette widths are not always the same. For example, if we consider passionflower, the most different sample employing HPLC data corresponded to an herbal drop preparation whereas the most different sample using UV–vis spectroscopic data was the pure herb. Similarly, the two considered lavender samples were found to be much closer using HPLC data than spectroscopic data. These facts point out that the information provided by each technique is not the same and may be complementary, which prompted the application of data fusion.

In order to build a model combining the information provided by chromatographic and UV–vis spectroscopic profiles, the scores obtained from the two PCA models previously constructed were assembled in a single matrix. This mid-level data fusion approach presents several advantages over the simple combination of both fingerprints (as in raw data) including the possibility to perform specific data pretreatment in each profile and the inclusion of a similar number of data points from each technique. The latter is important to ensure that both techniques have a similar weight in the joint PCA model. The scores diagram obtained with PCA (Fig. 3a) shows a clear clusterization between the four types of herbal medicines, which was confirmed through the Silhouette analysis (Fig. 3b). As it can be observed in Table 1, this PCA model based

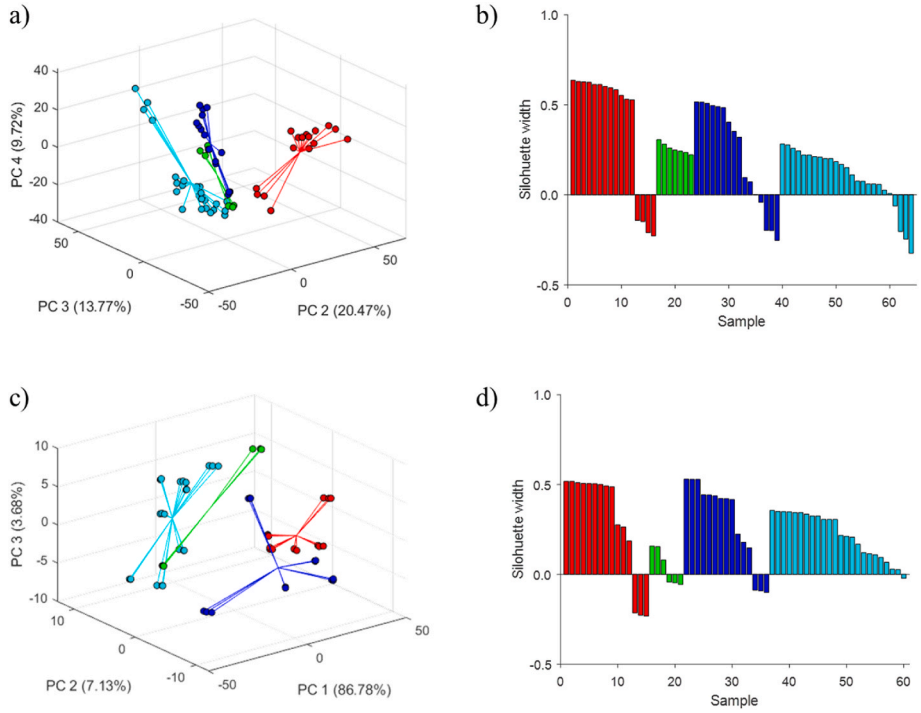

Fig. 2. Scores diagram (a, c) and Silhouette analysis (b, d) plots for PCA models constructed using individual data from chromatographic (a, b) and UV–vis spectroscopic (c, d) profiles. Colour code: chamomile (red), lavender (green), passionflower (blue) and valerian (cyan). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Percentage of samples with a positive and negative Silhouette width obtained from PCA models using chromatographic profiles, UV–vis spectroscopic profiles and the combination of both data sets.

		Chamomile	Lavender	Passionflower	Valerian	Global
HPLC data	Silhouette width >1	75	100	73	84	81
	Silhouette width <1	25	0	27	16	19
UV–vis data	Silhouette width >1	80	50	80	96	85
	Silhouette width <1	20	50	20	4	15
Combined data	Silhouette width >1	100	100	80	88	90
	Silhouette width <1	0	0	20	12	10

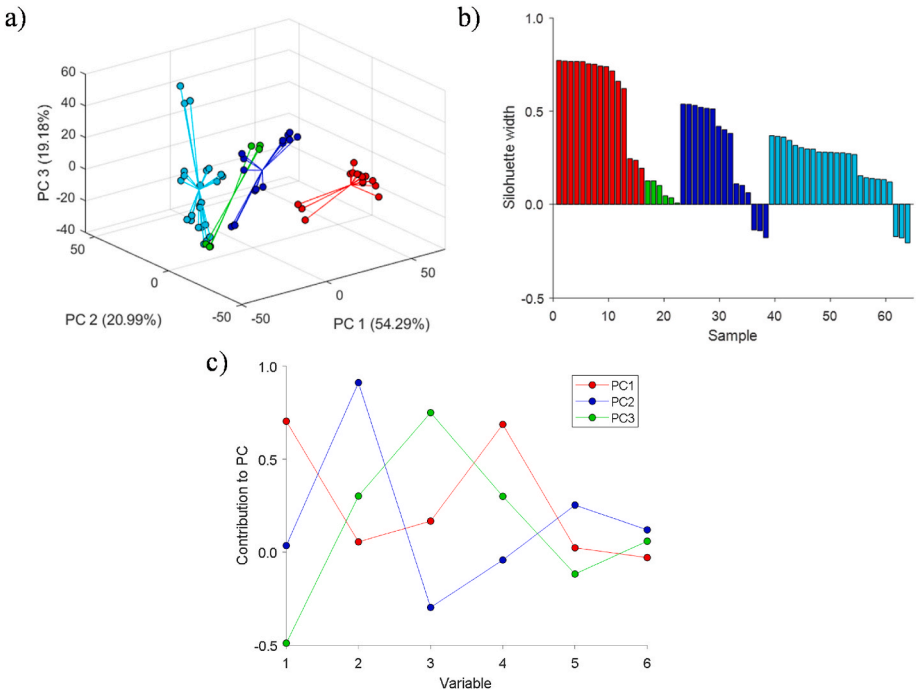

Fig. 3. Scores diagram (a), Silhouette analysis (b), and loadings (c) plots for PCA models constructed using joint data from chromatographic and UV–vis spectroscopic profiles. Colour code for (a) and (b): chamomile (red), lavender (green), passionflower (blue) and valerian (cyan) (c). In the loadings plot, variables 1 to 3 correspond to HPLC and variables 4 to 6 to UV–vis spectroscopic data. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

on joined information from chromatographic and spectroscopic profiles provided a better clusterization (90% of samples with positive Silhouette width) than the PCA models previously obtained with separate data. In particular, it should be stressed that this model was able to correctly clusterize all chamomile and lavender samples, providing negative Silhouette widths only for one sample of valerian and one sample of passionflower. The loadings of this PCA model (Fig. 3c) indicated that HPLC data had a higher contribution to the joint PCA model, as the scores on all three PCs deriving from the chromatographic PCA model (variables 1 to 3) had a great contribution to at least one PC. Nevertheless, the first variable related to spectroscopic data (variable 4) also contributed greatly to PC1 and, to a lesser extent, variable 5 contributed to PC2. This demonstrates that both analytical techniques provided complementary information, which benefits a data fusion approach.

The results provided in the exploratory analysis revealed that data fusion is the approach that has a better potential in differentiating among different classes of herbal herbs when several herbal preparation types are considered. Thus, the combined data from chromatographic and UV–vis spectroscopic profiles was considered for a further classification model, which was based on PLS-DA. For this purpose, the whole set of samples was distributed between a training (60% of samples) and a test set (40% of samples), and a PLS-DA model was built. It should be pointed out that, although lavender samples were included in the test set as out-of-class samples, the low number of available samples prevented the inclusion of lavender as a main class when building the PLS-DA model. In the developed PLS-DA model the optimum number of LVs was 3 (Fig. S1) and the cumulative variances explained in X and Y blocks were 92.9% and 51.6%, respectively. The model capability was evaluated by means of the receiver operating characteristics curves (ROC), which are displayed in Fig. 4. As it can be observed, the model provided an excellent discrimination capability for all three classes, with an area under the curve (AUC) of 1.000 (Fig. 4a–c). Discriminant thresholds were defined as the point where the specificity line crosses with the sensitivity line (Fig. 4d–f), thus minimizing the number of false positives and false negatives. As expected from the ROC curves, the selected threshold enabled the PLS-DA model to correctly predict the class of all samples in the training set (Fig. 5, Table S2). Consequently, ideal values were obtained for all classes in the training set, with low root mean square errors (RMSE), 100% non-error classification rate (%NER) and classification rate (%CCR), 1.000 values for both sensitivity and selectivity (Table 2), and an accuracy of 100%. The developed PLS-DA model was then externally validated using the test set, providing perfect predictions in the case of chamomile and passionflower and only misspredicting three replicates of one sample of lavender as valerian (Fig. 5, Table S2). RMSE values were still low for the test set and excellent discrimination figures of merit were obtained (Table 2), with a global accuracy of 87.5%. Additionally, the advantages of employing a data fusion approach were confirmed through the VIP scores (Fig. 6), which revealed that the most important variables (i.e. those with VIP score higher than 1) in the discrimination between chamomile, passionflower, and valerian were all three scores from HPLC profiles (variables 1 to 3 in Fig. 6) and the first score from spectroscopic profiles (variable 4). Moreover, although PLS-DA models built using solely optical or chromatographic profiles provided an excellent performance for

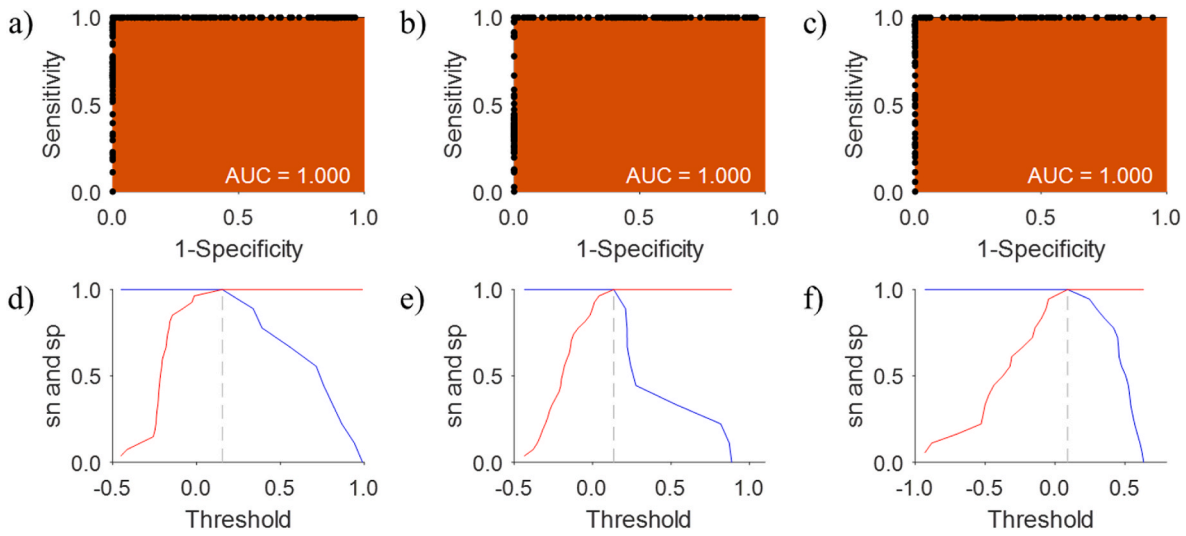

Fig. 4. ROC curves (a–c) and plots of sensitivity (sn, blue) and specificity (sp, red) values as the class threshold is changed (d–f), for each class considered in the PLS-DA model: chamomile (a,d), passionflower (b,e) and valerian (c,f). In d-f, the vertical grey dashed line represents the discriminant threshold: 0.153 for chamomile, 0.135 for passionflower and 0.091 for valerian. AUC: area under the curve. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

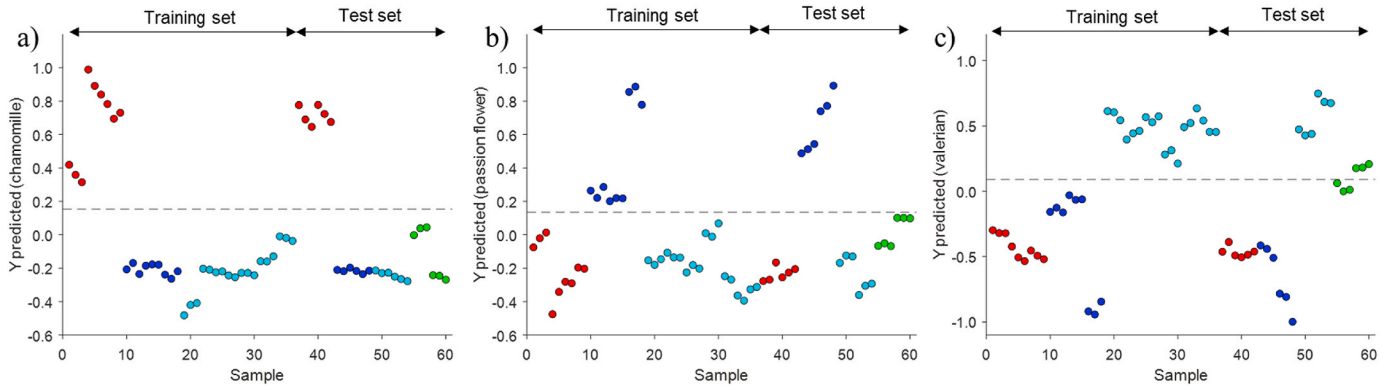

Fig. 5. Calculated response obtained for each class in the developed PLS-DA model for both training and test sets: chamomile (a), passionflower (b) and valerian (c). The horizontal grey dashed line represents the discriminant threshold: 0.153 for chamomile, 0.135 for passionflower and 0.091 for valerian. Colour code: chamomile (red), lavender (green), passionflower (blue) and valerian (cyan). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Classification parameters (root mean square error (RMSE), non-error classification rate (%NER), classification rate (%), class specificity, and sensitivity) obtained in training and test sets. For training set, parameters were calculated by cross-validation based on venetian blinds, using 10 splits and 1 sample per split.

	PLS-DA model	RMSE	%NER	Classification rate (%)	Specificity	Sensitivity
Training set	Chamomile	0.3100	100.0	100.0	1.000	1.000
	Passionflower	0.3925	100.0	100.0	1.000	1.000
	Valerian	0.5486	100.0	100.0	1.000	1.000
	Average		**100**	**100**	**1.000**	**1.000**
Test set	Chamomile	0.2517	100.0	100.0	1.000	1.000
	Passionflower	0.2928	100.0	100.0	1.000	1.000
	Valerian	0.5473	91.7	87.5	0.833	1.000
	Average		**97.2**	**95.8**	**0.944**	**1.000**

the training set (100% accuracy in both cases), the prediction of the test set was not as good, yielding an accuracy of 75% for both optical and chromatographic data, which is much lower than the 87.5% provided by the PLS-DA model built with the data fusion approach. Thus, the obtained results fully prove that the proposed UV–vis spectroscopic and chromatographic data fusion strategy would be an excellent tool for the authentication of herbal medicines.

4. Conclusions

The results obtained in this work demonstrate the benefits of data fusion in the discrimination of complex systems such as herbal medicines, in which great variety can be found in a single class, for example, among different types of herbal preparations. For this purpose, chromatographic and UV–vis spectroscopic profiles were first assessed separately, confirming through PCA and Silhouette analysis that the

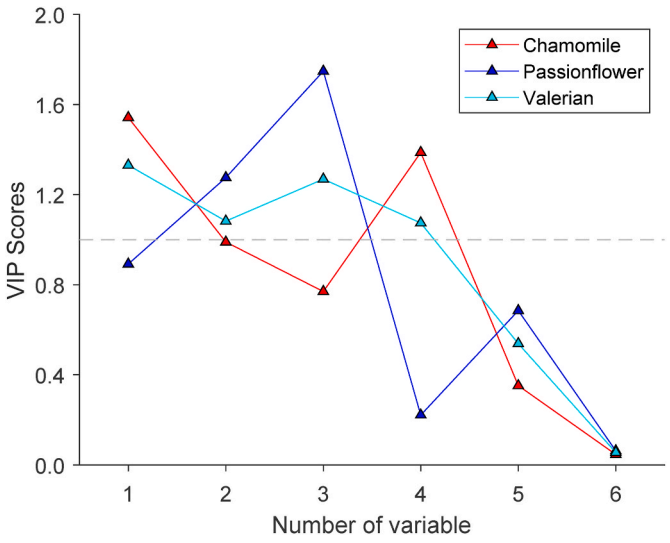

Fig. 6. VIP scores plots for the PLS-DA model constructed using joint data from chromatographic and UV–vis spectroscopic profiles. In the VIP scores plot, variables 1 to 3 correspond to HPLC and variables 4 to 6 to UV–vis spectroscopic data.

experimental profiles acquired were appropriate for the discrimination of chamomile, lavender, passionflower and valerian (81 and 85% of samples presented a positive Silhouette width using chromatographic and spectroscopic data, respectively). Nevertheless, the problematic samples were not the same in both models, which suggested that they were not relying on the same type of chemical information and, thus, data fusion could be a more suitable approach. Indeed, Silhouette analysis on the PCA model constructed with joint data showed an improved clusterization, with 90% of samples presenting a positive Silhouette width. Thus, data fusion approach was selected as the best option for herbal medicine discrimination and a PLS-DA model was built considering chamomile, passionflower and valerian as classes. The developed model was able to perfectly discriminate between the three classes of soothing herbs, overcoming the difficulties encountered when dealing with different types of herbal preparations.

Credit author statement

Clara Pérez-Ràfols: Conceptualization; Methodology; Formal analysis; Writing - original draft; Visualization.

Núria Serrano: Conceptualization; Methodology; Formal analysis; Writing - original draft; Visualization.

José Manuel Díaz-Cruz: Methodology; Validation; Visualization; Writing - review & editing; Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by PID2019-107102RB-C22 funded by MCIN/AEI/10.13039/501100011033 and the Generalitat de Catalunya (Project 2021 SGR 00006).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2023.104783.

References

[1] WHO, Guidelines for selecting marker substances of herbal origin for quality control of herbal medicines - TRS 1003 - Annex 1, Annex 1, 1003, WHO Technical Series, 2017.
[2] B.B. Petrovska, Historical review of medicinal plants' usage, Pharm. Rev. 6 (2012) 1–5, https://doi.org/10.4103/0973-7847.95849.
[3] A. Sharma, P. Sabharwal, R. Dada, Herbal medicine — An introduction to its history, in: R. Henkel, A. Agarwal (Eds.), Herbal Medicine in Andrology, Elsevier, 2021, pp. 1–8, https://doi.org/10.1016/b978-0-12-815565-3.00001-1.
[4] J.K. Patra, G. Das, S. Lee, S.S. Kang, H.S. Shin, Selected commercial plants: a review of extraction and isolation of bioactive compounds and their pharmacological market value, Trends Food Sci. Technol. 82 (2018) 89–109, https://doi.org/10.1016/j.tifs.2018.10.001.
[5] R. Gao, Y. Hu, Y. Dan, L. Hao, X. Liu, J. Song, Chinese herbal medicine resources: where we stand, Chin Herb Med 12 (2020) 3–13, https://doi.org/10.1016/j.chmed.2019.08.004.
[6] TechNavio, Global Herbal Medicine Market 2022-2026, 2021. ID: 5515025.
[7] A.H. Alostad, D.T. Steinke, E.I. Schafheutle, Herbal medicine classification: policy recommendations, Front. Med. 7 (2020) 31, https://doi.org/10.3389/fmed.2020.00031.
[8] J.K. Srivastava, E. Shankar, S. Gupta, Chamomile: a herbal medicine of the past with a bright future (review), Mol. Med. Rep. 3 (2010) 895–901, https://doi.org/10.3892/mmr.2010.377.
[9] D. Benke, A. Barberis, S. Kopp, K.H. Altmann, M. Schubiger, K.E. Vogt, U. Rudolph, H. Möhler, GABAA receptors as in vivo substrate for the anxiolytic action of valerenic acid, a major constituent of valerian root extracts, Neuropharmacology 56 (2009) 174–181, https://doi.org/10.1016/j.neuropharm.2008.06.013.
[10] S. Akhondzadeh, H.R. Naghavi, M. Vazirian, A.S. Pharmd, H. Rashidi Pharmd, M. Khani, Passionflower in the treatment of generalized anxiety: a pilot double-blind randomized controlled trial with oxazepam, J. Clin. Pharm. Therapeut. 26 (2001) 363–367, https://doi.org/10.1046/j.1365-2710.2001.00367.x.
[11] P.H. Koulivand, M. Khaleghi Ghadiri, A. Gorji, Lavender and the nervous system, evidence-based complementary and alternative medicine, 2013, 2013, https://doi.org/10.1155/2013/681304.
[12] M. Goodarzi, P.J. Russell, Y. vander Heyden, Similarity analyses of chromatographic herbal fingerprints: a review, Anal. Chim. Acta 804 (2013) 16–28, https://doi.org/10.1016/j.aca.2013.09.017.
[13] X. miao Liang, Y. Jin, Y. ping Wang, G. wa Jin, Q. Fu, Y. sheng Xiao, Qualitative and quantitative analysis in quality control of traditional Chinese medicines, J. Chromatogr. A 1216 (2009) 2033–2044, https://doi.org/10.1016/j.chroma.2008.07.026.
[14] P. Xie, S. Chen, Y. zeng Liang, X. Wang, R. Tian, R. Upton, Chromatographic fingerprint analysis-a rational approach for quality assessment of traditional Chinese herbal medicine, J. Chromatogr. A 1112 (2006) 171–180, https://doi.org/10.1016/j.chroma.2005.12.091.
[15] P. Wang, L. Li, H. Yang, S. Cheng, Y. Zeng, L. Nie, H. Zang, Chromatographic fingerprinting and quantitative analysis for the quality evaluation of Xinkeshu tablet, J Pharm Anal 2 (2012) 422–430, https://doi.org/10.1016/j.jpha.2012.05.006.
[16] World Health Organization, Guidelines for the assessment of herbal medicines, 1991. Munich.
[17] US Food and Drug Administration, FDA Guidance for industry — Botanical drug products (draft guidance), 2000. Rockville.
[18] S.C.A. Cobzac, D. Casoni, M. Babea, B. Balabanova, N.M. Ruzdik, Ultraviolet-visible (UV-vis) spectroscopy and cluster analysis as a rapid tool for classification of medicinal plants, Stud. Univ. Babes-Bolyai, Chem. 64 (2019) 191–203, https://doi.org/10.24193/subbchem.2019.4.14.
[19] Y. Li, Y. Shen, C. liang Yao, D. an Guo, Quality assessment of herbal medicines based on chemical fingerprints combined with chemometrics approach: a review, J. Pharm. Biomed. Anal. 185 (2020), 113215, https://doi.org/10.1016/j.jpba.2020.113215.
[20] F. Castanedo, A review of data fusion techniques, Sci. World J. (2013), https://doi.org/10.1155/2013/704504, 2013.
[21] S.M. Azcarate, R. Ríos-Reina, J.M. Amigo, H.C. Goicoechea, Data handling in data fusion: methodologies and applications, Trends Anal. Chem. 143 (2021), 116355, https://doi.org/10.1016/j.trac.2021.116355.
[22] X. Cetó, N. Serrano, M. Aragó, A. Gámez, M. Esteban, J.M. Díaz-Cruz, O. Núñez, Determination of HPLC-UV fingerprints of Spanish paprika (Capsicum annuum L.) for its classification by linear discriminant analysis, Sensors (2018) 18, https://doi.org/10.3390/s18124479.
[23] Matlab, Mathworks Inc, Version R2021b, 2021. Natick, MA, USA.
[24] PLS_Toolbox, Eigenvector Research, Inc, Version 8.9.2, 2021. Manson, WA USA 98831.
[25] J.A. Navarro-Huerta, J.R. Torres-Lapasió, S. López-Ureña, M.C. García-Alvarez-Coque, Assisted baseline subtraction in complex chromatograms using the BEADS algorithm, J. Chromatogr. A 1507 (2017) 1–10, https://doi.org/10.1016/j.chroma.2017.05.057.

[26] X. Ning, I.W. Selesnick, L. Duval, Chromatogram baseline estimation and denoising using sparsity (BEADS), Chemometr. Intell. Lab. Syst. 139 (2014) 156–167, https://doi.org/10.1016/j.chemolab.2014.09.014.
[27] S.J. Dixon, N. Heinrich, M. Holmboe, M.L. Schaefer, R.R. Reed, J. Trevejo, R. G. Brereton, Use of cluster separation indices and the influence of outliers: application of two new separation indices, the modified silhouette index and the overlap coefficient to simulated data and mouse urine metabolomic profiles, J. Chemom. 23 (2009) 19–31, https://doi.org/10.1002/cem.1189.
[28] D. Seifu, F. Assefa Zewdie, S.M. Abay, F. Assefa, Medicinal plants as antioxidant agents: Understanding their mechanism of action and therapeutic efficacy, Trams world research network, 2012.