Macedonian Pharmaceutical Bulletin, 72 (Suppl 1) 403 - 404 (2026)
Online ISSN: 1857-8969
DOI: 10.33320/maced.pharm.bull.2026.72.03.194
Short communication

Data-driven modeling and green analytics of medicinal Sage and related *Salvia* species

Ivana Cvetkovikj Karanfilova[1*], Maja Simonoska Crcarevska[2], Veronika Stoilkovska Gjorgievska[1], Ana Trajkovska Matevska[1], Iskra Davkova[1], Filip Todorov[1], Marija Karapandzova[1], Gjoshe Stefkov[1]

[1]*Institute of Pharmacognosy, Faculty of Pharmacy, Ss. Cyril and Methodius University in Skopje, Majka Tereza 47, 1000 Skopje, Republic of North Macedonia*
[2]*Institute of Pharmaceutical Technology, Faculty of Pharmacy, Ss. Cyril and Methodius University in Skopje, Majka Tereza 47, 1000 Skopje, Republic of North Macedonia*

Introduction

The genus *Salvia* (Lamiaceae) comprises more than 1000 species and represents one of the most important groups of medicinal and aromatic plants. Several species, including *S. officinalis*, *S. fruticosa* and *S. pomifera*, are recognized as valuable sources of essential oils and polyphenolic compounds associated with antioxidant, antimicrobial, anti-inflammatory, antidiabetic, and neuroprotective activities (Cvetkovikj et al., 2013; 2015; 2016).

In recent years, the field of medicinal plant analysis has undergone a significant transformation driven by green analytical chemistry principles, high-throughput chromatographic technologies, and advanced data-driven approaches. In particular, machine learning-compatible analytical frameworks have enabled more comprehensive interpretation of complex phytochemical datasets, facilitating species discrimination, chemotype classification, and quality assessment (Abraham and Kellogg, 2021; Ichim and Booker, 2021).

Within this context, large-scale previously generated datasets from multiple *Salvia* populations provide an excellent model for integrating classical pharmacognostic evaluation with modern data-driven approaches, enabling more efficient, sustainable, and reproducible plant characterization.

Materials and methods

Plant material and analytical workflow

Previously published phytochemical datasets, supplemented with newly generated phytochemical data, covering more than 65 *Salvia* populations, including *S. officinalis* (n > 50), *S. fruticosa* (n > 15), and S. *pomifera* (n = 2), originating from 10 Southeast European countries, were used for integrative data-driven evaluation. The datasets included essential oil profiles obtained by hydrodistillation and GC-MS (Cvetkovikj et al., 2015), HS-GC/MS volatile profiles obtained directly from fresh and dried sage leaves (Cvetkovikj et al., 2016), and polyphenolic profiles obtained after ultrasound-assisted extraction and LC/DAD/ESI-MS^n analysis using core-shell column technology (Cvetkovikj et al., 2013).

Principal Component Analysis (PCA), Partial Least Squares (PLS) regression and hierarchical clustering analysis (HCA) were applied to evaluate phytochemical variability and identify marker compounds responsible for species differentiation. In parallel, the datasets were organized into a

*ivanacvetkovikj@ff.ukim.edu.mk

structured, machine-learning compatible format suitable for further unsupervised and supervised classification of *Salvia* species and identification of phytochemical variables with potential predictive value.

Results and discussion

The previously generated phytochemical datasets revealed considerable variability in essential oil and polyphenolic composition among the investigated *Salvia* populations. A total of 122 volatile constituents and 53 polyphenolic compounds belonging to different chemical classes were previously identified (Cvetkovikj et al., 2013; 2015; 2016). The present data-driven evaluation enabled clearer discrimination of the investigated taxa according to their characteristic phytochemical profiles and highlighted compounds contributing to species differentiation.

From the green analytical perspective, HS-GC/MS enabled rapid screening of volatile fractions directly from plant material without solvent consumption, whereas ultrasound-assisted extraction and core-shell LC technology reduced extraction and analysis time (Cvetkovikj et al., 2013; 2016). These workflows generated robust phytochemical fingerprints suitable for large-scale quality assessment and botanical authentication.

PLS modelling demonstrated that selected essential oil constituents, including 1,8-cineole, camphor, camphene, borneol acetate and α-humulene, contributed significantly to the differentiation of *Salvia* species and reflected both taxonomic and environmental influences. In contrast, several polyphenolic markers, particularly luteolin diglucuronide, luteolin 7-O-glucuronide and a caffeic acid derivative, were found to be more species-dependent and less affected by climatic factors, highlighting their potential as stable chemotaxonomic markers.

Hierarchical clustering analysis further confirmed the discrimination of the investigated taxa, separating *S. officinalis* and *S. fruticosa* into distinct clusters, while revealing characteristic phytochemical signatures associated with each species.

Conclusion

The integration of green extraction procedures, high-throughput chromatographic techniques, and data-driven approaches enables comprehensive and efficient phytochemical characterization of *Salvia* species while minimizing solvent consumption, analysis time, and environmental impact. The large-scale phytochemical dataset generated from Balkan *Salvia* populations provides a robust foundation for data-driven modelling and classification approaches, supporting reliable species discrimination, chemotype evaluation, and botanical authentication. These findings highlight the value of combining sustainable analytical strategies with machine-learning-compatible data frameworks, supporting the transition from conventional phytochemical characterization toward data-driven pharmacognostic evaluation and botanical authentication.

References

Cvetkovikj, I., et al., 2013. Polyphenolic characterization and chromatographic methods for fast assessment of culinary *Salvia* species from South East Europe. J. Chromatogr. A 1282, 38–45. https://doi.org/10.1016/j.chroma.2013.01.045

Cvetkovikj, I., et al., 2015. Essential oils and chemical diversity of southeast European populations of *Salvia officinalis* L. Chem. Biodivers.12(7), 1025-1039. https://doi.org/10.1002/cbdv.201400273

Cvetkovikj, I., et al., 2016. Headspace screening: A novel approach for fast quality assessment of the essential oil from culinary sage. Food Chem. 202, 133-140. https://doi.org/10.1016/j.foodchem.2016.01.141.

Abraham, E.J., and Kellogg, J.J., 2021. Chemometric-Guided Approaches for Profiling and Authenticating Botanical Materials. Front. Nutr. 8:780228. doi: 10.3389/fnut.2021.780228

Ichim, M.C., and Booker, A., 2021. Chemical Authentication of Botanical Ingredients: A Review of Commercial Herbal Products. Front. Pharmacol. 12:666850. doi: 10.3389/fphar.2021.666850