

Machine learning algorithms at the bright side of FTIR-ATR analysis in *Lathyrus sativus* L

Elsa Mecha^{1,2*}; Pedro R. Brandão^{1,3}; Letice Gonçalves²; Elsa Brito¹; Andreia Bento da Silva⁴; Maria Carlota Vaz Patto²; Maria Rosário Bronze^{1,2,4}

¹Instituto de Biologia Experimental e Tecnológica, Oeiras, Portugal; ²Instituto de Tecnologia Química e Biológica António Xavier, Oeiras, Portugal; ³LAQV-REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, FCT NOVA, Universidade NOVA de Lisboa, Caparica, Portugal; ⁴Faculdade de Farmácia, Universidade de Lisboa, Lisboa, Portugal

*emecha@ibet.pt; emecha@itqb.unl.pt

DOI: DOI: 10.5281/zenodo.15262025

Keywords: grass pea; beta-ODAP; homoarginine; spectroscopy; mass spectrophotometry; legumes; quality

Lathyrus sativus L., also called grass pea, possesses favourable nutritional characteristics as well as adaptative traits for handling stress [1]. Its consumption associated to a very restrictive non-diversified diet, has been linked to the onset of lathyrism. The discovery of a neurotoxin β -N-oxalyl-L- α , β -diaminopropionic acid (β -ODAP), tarnished the image of grass pea as a valuable food source [2]. Nevertheless, grass pea has undeniable nutritional value, being a precious source of L-homoarginine, L-hArg, a promising cardio protector compound [3].

Advanced vibrational spectroscopy to characterize legumes nutritional quality has been applied for different legumes [4,5], but its use to characterize grass pea accessions is still very limited. In this study, we applied ATR-FTIR spectroscopy supported by machine learning methods focusing on the development of robust, and cost-effective models for ODAP, homoarginine and phenolic compounds content in grass pea.

For that, we cropped more than 200 different accessions of grass pea over three years under similar conditions in Alvaiázere (Portugal). The aqueous extracts were analysed by HPLC-ESI-MS/MS to determine ODAP and homoarginine contents [6]. The ethanolic extracts were analysed by spectrophotometry to determine the total phenolic content by Folin-Ciocalteu's method, and the *in vitro* antioxidant activity by Oxygen radical absorbance capacity (ORAC) [7]. We screened the samples applying the Attenuated Total Reflectance – Fourier Transformed Infrared Spectroscopy, ATR-FTIR (Thermo Nicolet 6700 spectrometer).

Different algorithms were applied to more than 2000 collected spectra based on partial least square regression multivariate analysis, kernelized Support Vector Machine (SVMs), 1D convolutional neural nets e random forests, combined with variable selection methods to restrict the spectrum to relevant areas were applied using Python interface programming language to recognize patterns and explore the potential correlations between ATR-FTIR spectral data and the quantified quality parameters.

Following the numerous combinations of pre-processing methods and hyperparameters, 9600 pipelines were tuned by random search cross-validation. Data suggested the adequacy of ATR-FTIR for predicting quality parameters only after specific pre-processing methods (e.g., detrend, *Savitzky-Golay* derivatives, and multiplicative scatter correction, MSC). ATR-FTIR spectroscopy was able to derive alpha-ODAP ($R^2=0.73$), beta-ODAP ($R^2=0.86$), and Homoarginine ($R^2=0.78$), considering the spectral region between 1806 and 892 cm^{-1} . Other algorithms are currently being tested.

The results of the present work suggest that ATR-FTIR spectral data can represent fundamental tools to support future breeding programs dedicated to the improvement of grass pea's quality.

References:

- [1] Ramya KR, Tripathi K, Pandey A, Barpete S, Gore PG, Raina AP, Khawar KM, Swain N & Sarker A (2022) Rediscovering the potential of multifaceted orphan legume grass pea - a sustainable resource with high nutritional values. *Front Nutr* 8:826208. Doi: 10.3389/fnut.2021.826208.
- [2] Lambein F, Travella S, Kuo Y-H, Van MM, & Heijde M (2019). Grass pea (*Lathyrus sativus* L.): orphan crop, nutraceutical or just plain food?. *Planta* 250. Doi: 10.1007/s00425-018-03084-0.
- [3] Koch V, Gruenewald LD, Gruber-Rouh T, et al. (2023). Homoarginine in the cardiovascular system: Pathophysiology and recent developments. *Fundam Clin Pharmacol* 37(3): 519-529. Doi:10.1111/fcp.12858.
- [4] Johnson JB, Walsh K, Naiker M (2020). Application of infrared spectroscopy for the prediction of nutritional content and quality assessment of faba bean (*Vicia faba* L.). *Legume Sci* 2:e40. Doi: 10.1002/leg3.40.

- [5] Dıblan S, Kadiroğlu P, Aydemir LY (2018). FT-IR spectroscopy characterization and chemometric evaluation of legumes extracted with different solvents. *Food and Health* 4(2): 80-88. Doi: 10.3153/FH18008.
- [6] Bento-Silva A, Gonçalves L, Mecha E, Pereira F, Vaz Patto MC, Bronze MDR (2019). An improved HILIC HPLC-MS/MS Method for the determination of β -ODAP and its α isomer in *Lathyrus sativus*. *Molecules* 24(17):3043. Doi: 10.3390/molecules24173043.
- [7] Mecha E, Leitão ST, Carbas B, Serra AT, Moreira PM, Veloso MM, Gomes R, Figueira ME, Brites C, Vaz Patto MC, Bronze MR (2019). Characterization of soaking process' impact in common beans phenolic composition: contribute from the unexplored Portuguese germplasm. *Foods* 28; 8(8):296. Doi: 10.3390/foods8080296.

Acknowledgments:

FCT for funding QuaLaty (PTDC/AGR-TEC/0992/2014), and the Individual Call to Scientific Employment Stimulus - 6th Edition, Reference 2023. 07800.CEECIND. H2020 programme for funding DIVINFOOD (Grant Agreement nº 101000383).