Classification of apple and pear species from Alcobaça region (Portugal) and their cultivars with machine learning algorithms

João David Teixeira[1,2,3,*], **Claudia Sánchez**[4], **Carina Almeida**[1,5,6], **Ana Sanches-Silva**[1,3,7,8], **Pier Parpot**[2,9]

[1]National Institute for Agrarian and Veterinary Research (INIAV), I.P., Vairão campus, Rua dos Lágidos, Lugar da Madalena, Vila do Conde, Portugal
[2]University of Minho, Center of Chemistry, Braga, Portugal
[3]Center for Study in Animal Science (CECA), ICETA, University of Porto, Porto, Portugal
[4]National Institute for Agrarian and Veterinary Research (INIAV), I.P., Alcobaça campus, Alcobaça, Portugal
[5]LEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Rua Dr.Roberto Frias, 4200-465 Porto, Portugal
[6]AliCE – Associate Laboratory in Chemical Engineering, Faculty of Engineering, University of Porto, Rua Dr.Roberto Frias, 4200-465 Porto, Portugal
[7]University of Coimbra, Faculty of Pharmacy, Polo III, Azinhaga de Stª Comba, Coimbra, Portugal
[8]Associate Laboratory for Animal and Veterinary Sciences (Al4AnimalS), Lisbon, Portugal
[9]University of Minho, Center of Biological Chemistry, Braga, Portugal

e-mail: david.teixeira @iniav.pt

Principal Component Analysis (PCA) transforms the original variable into new ones called principal components (PC). These PC´s are calculated attributing a coefficient for each original variables proportional to their contribution into this transformation in order to maximize the variances of the first few components [1]. The main objective is to reduce the dimensionality, while keeping the contribution of all initial variables in order to provide a visual pattern recognition [2]. PCA biplot graphs with both scores and loadings provide information on the influence of each variable on a given sample. The hierarchical clustering was also employed in order to highlight the similarities among samples.

The following variables were determined through ethanolic extracts of apple and pear varieties from the Alcobaça region (Portugal) using spectrophotometric analysis: DPPH radical scavenging, β-carotene bleaching, total phenolic content, total flavonoid content, and fructose content. The results show that a very high correlation exists among the variables antioxidant capacity through DPPH, total phenolics content and total flavonoids, while the fructose content shows independent behavior in relation to the other ones. The graph of scores for the first two PCs, which explain 90% of variance, shows three different clusters with different apple and pear species. It can be concluded from these results that the fructose content allows to separate apples and pears while antioxidant capacity through DPPH, total phenolics content and total flavonoids content can be used to separate different cultivars of each fruit.

This study shows that multivariate analysis, with special focus on PCA, can be a valuable tool for the separation of different fruit species and their cultivars highlighting the similarities and differences among them.

References:

[1] L. A. Berrueta, R. M. Alonso-Salces, e K. Héberger, Journal of Chromatography A (2007), 196

[2] M.Á. Rodríguez-Delgado, G. González-Hernández, J.E. Conde-González, J.P. Pérez-Trujillo, Food Chemistry (2002), 523