

EUROSENSORS 2014, the XXVIII edition of the conference series

Electrochemical multi-sensors device coupled with heuristic or meta-heuristic selection algorithms for single-cultivar olive oil classification

António M. Peres^{a*}, Ana C.A. Veloso^{b,c}, José A. Pereira^d, Luís G. Dias^d

^aLSRE, ESA, Instituto Politécnico de Bragança, Campus Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal

^bInstituto Politécnico de Coimbra, ISEC, DEQB, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal

^cCEB, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

^dCIMO, ESA, Instituto Politécnico de Bragança, Campus Santa Apolónia, Apartado 1172, 5301-855 Bragança, Portugal

Abstract

Potentiometric electrochemical multi-sensors' performance highly depends on the capability of selecting the best set of sensors. Indeed, signals are usually collinear resulting in over-fitted multivariate models with low predictive applicability. In this work, a comparative study was made to evaluate the predictive performance of classification models coupled with heuristic or meta-heuristic variable selection algorithms. In this study, eleven single-cultivar extra virgin olive oils, from two crop years, were used. The results demonstrated that linear discriminant analysis with simulated annealing algorithm allowed selecting the best subset of sensors enabling 100% of correct cross-validation classifications, considering samples split by crop year.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Peer-review under responsibility of the scientific committee of Eurosenors 2014

Keywords: Potentiometric electrochemical multi-sensors; linear discriminant analysis; heuristic selection algorithms; meta-heuristic selection algorithms; single-cultivar extra virgin olive oils

1. Introduction

The design of electrochemical multi-sensors devices requires the selection of sensors that maximize their capability to provide a useful chemical fingerprint of the samples [1]. The complexity of the signals arising from the electrochemical devices requires the application of statistical tools for reducing the number of sensors to improve the

* Corresponding author. Tel.: +351-273-303-220; fax: +351-273-332-405.

E-mail address: peres@ipb.pt

multivariate data processing by eliminating the presence of irrelevant information, overcoming or minimizing co-linearity between the variables, reducing random noise or unwanted systematic variations and keeping limited the number of needed samples to implement a robust predictive model [1-3]. However, usual variable selection approaches are not always capable of selecting the most informative predictors to build simple multivariate linear models and, in several cases, less straightforward and highly complex non-linear models must be used.

In this work, a comparative evaluation study between heuristic and meta-heuristic variable selection algorithms is done to reduce the number of sensors that must be included in a linear discriminant analysis (LDA) model for classifying extra virgin olive oils (EVOOs) according to their cultivars and crop year. So, the practical aim was to demonstrate the potential of a potentiometric multi-sensor device coupled with the best selection approach to discriminate eleven single-cultivar Spanish EVOOs. These EVOOs are highly appreciated and an important component of the Mediterranean diet [4], being prone to frauds involving mislabelling and adulteration [5].

2. Materials and methods

2.1. Extra-virgin olive oil samples and extraction procedure

Spanish EVOOs of eleven different olive cultivars (cvs. Arbequina, ARB; Arbosana, ARBO; Arroniz, ARR; Cornicabra, COR; Frantoio, FRA; Hojiblanca, HOJ; Manzanilla, MAN; Picual, PIC; Redondilla, RED; Royela, ROY; and Zorzal, ZOR) were obtained directly from certified olive oil producers of the north of Spain (Valladolid region) during two production years (2011 and 2012). As previously referred [5], all olive oil samples analyzed were packed and stored in the dark at $-20\text{ }^{\circ}\text{C}$ in a 24-h period after their production and kept under those conditions until further analysis. The EVOO samples were subjected to extraction using a hydroethanolic solution ($\text{H}_2\text{O}:\text{EtOH}$, 80:20 v/v), to obtain a polar compounds rich-solution, as described in detail by Dias et al. [5]. Two independent samples were collected from each monovarietal EVOO and analyzed electrochemically in duplicate.

2.2. Potentiometric multi-sensor device and analysis

The electrochemical multi-sensor device used (an electronic tongue, Fig. 1) includes two print-screen devices, each composed by an array with the same 20 potentiometric sensors, based on all-solid-state potentiometric electrodes with polymeric membranes. These chemical sensors have high stability and cross sensitivity to different species in solution (non-specific and with low-selectivity polymeric lipidic membranes) and are suitable for analyzing complex liquid samples as previously reported [5]. Each sensor was identified by a code using the letter S (for sensor), followed by the number of the array (1 or 2), followed by the number of the polymeric membrane (1–20, where each number represents a different mixture of plasticizer and additive used). The potentiometric signal from each sensor (varying from -1.0 to $+1.0$ V) was measured against an Ag/AgCl reference electrode and recorded using a multiplexer equipment from Agilent.



Fig. 1. Potentiometric multi-sensor system built using print-screen technique with 20 polymeric membranes (in duplicate).

2.3. Statistical analysis

All statistical analysis were performed using the *klaR* [6], *Subselect* [7,8] and *MASS* [9] packages of the open source statistical program R (version 2.15.1).

For classification purposes a supervised LDA was applied the potentiometric sensor signals recorded. A leave-one-out cross-validation (LOO-CV) procedure was used to minimize over-fitting and consequently overoptimistic results. Subsets of the most informative independent variables (containing 2 to 35 sensors of the 40 potentiometric sensors used in the E-tongue) were selected using common heuristic methods (sequential forward selection, SFS; sequential backward (elimination) selection, SBS; and the stepwise selection, STS) and meta-heuristic algorithm (simulated annealing, SA) [10,11]. The purpose was to verify the potential of a variable selection algorithm in establishing the best sensors subset that enabled maximizing the predictive correct classification performance with the minimum number of sensors in the multivariate model.

3. Results

In a previous work, Dias et al. [5] demonstrated that a potentiometric E-tongue could be successfully (60 to 100% of correct LOO-CV classifications) applied to discriminate monovarietal EVOOs (from two crop years) according to olive cultivar, when only 3-4 single-cultivar EVOOs were considered. Indeed, when six monovarietal EVOOs were evaluated simultaneously the classification performance decreased to 43%. This result can be explained since olive oils of the same olive cultivar have been produced in different years (2011 and 2012) and so, influenced by different edaphoclimatic conditions and storage time, probably affecting their organoleptic characteristics (like bitterness, astringency or pungency), which are related with the polar compounds content of the hydro-ethanolic extracts analyzed with the E-tongue. Therefore, in this study, samples were split according to EVOO's production year, resulting in 9 different single-cultivar EVOOs per year:

- 2011 (22 samples × 2 extractions) – cvs. ARB, ARBO, ARR, COR, FRA, HOJ, MAN, PIC and RED.
- 2012 (22 samples × 2 extractions) – cvs. ARB, ARBO, COR, FRA, HOJ, PIC, RED, ROY and ZOR.

For each year, the classification predictive linear discriminant models were established using heuristic and meta-heuristic variable selection algorithms. The results showed that heuristic techniques failed in selecting an adequate sensors subset allowing a predictive LOO-CV correct classification lower than 53% (Fig. 2). This drawback could be due to difficulties on dealing with the collinear potentiometric signal profiles, not allowing selecting multivariate classification models containing more than 17 sensors. On the contrary, SA algorithm enabled establishing linear discriminant models for sensors subsets containing 2 to 35 sensors. The best LDA-SA models, one for each year, were established base on 29 potentiometric sensor signals recorded in the analysis of EVOOs hydro-ethanolic extracts. For production years 2011 and 2012, the models enabled 100% of correct classifications for original grouped data (Fig. 3), and 93.2% (2011) and 95.5% (2012) predictive LOO-CV correct classifications.

4. Conclusions

The work described showed that a potentiometric E-tongue coupled with a LDA-SA procedure could be used as a fast, cost-effective tool for single-cultivar EVOOs classification. Indeed, this approach may be foreseen as a possible practical analytical methodology for guarantying cultivar authentication in single-cultivar EVOOs, minimizing the risk of fraud of this high price food product. Nevertheless, the proposed methodology requires assembling olive oils according to their production year, minimizing the differences in EVOOs organoleptic characteristics due to the influence of the edaphoclimatic conditions observed in year.

Acknowledgements

This work was partially co-financed by FCT and FEDER under Program COMPETE (Project PESt-C/EQB/LA0020/2013); by the Strategic Project PESt-OE/EQB/LA0023/2013 and by the project ref. RECI/BBB-EBI/0179/2012 (project number FCOMP-01-0124-FEDER-027462) funded by Fundação para a Ciência e a Tecnologia.

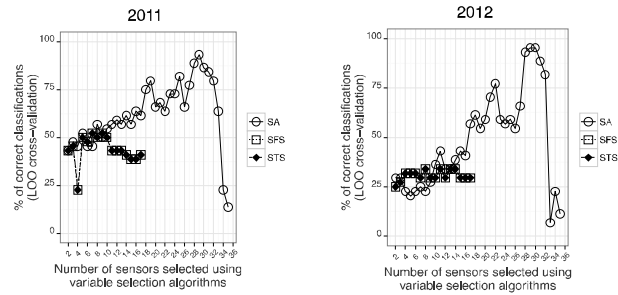


Fig. 2. Correct LOO cross-validation classification percentages versus the number of sensors in the single-cultivar EVOOs linear discriminant analysis models for each crop year, selected with simulated annealing (SA), sequential forward selection (SFS) and stepwise selection (STS) algorithms.

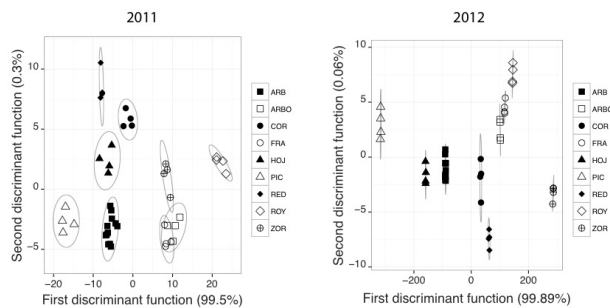


Fig. 3. Linear discriminant analysis of eleven single-cultivars EVOOs using the best model achieved with simulated annealing algorithms for each crop year. The ellipses in each group are associated with the confidence interval of 95%, assuming a multivariate normal distribution.

References

[1] J.M. Gutiérrez, Z. Haddi, A. Amari, B. Bouchikhi, A. Mimendia, X. Cetó and M. del Valle, Hybrid electronic tongue based on multisensor data fusion for discrimination of beers, *Sens. Actuat. B-Chem.*, 177 (2013) 989–996.

[2] N. Prieto, P. Oliveri, R. Leardi, M. Gay, C. Apetrei, M.L. Rodríguez-Méndez and J.A. Saja, Application of a GA-PLS strategy for variable reduction of electronic tongue signals, *Sens. Actuat. B-Chem.*, 183 (2013) 52–57.

[3] M. Liu, M. Wang, J. Wang and D. Li, Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar, *Sens. Actuat. B-Chem.*, 177 (2013) 970–980.

[4] C. Ruiz-Samblás, A. Tres, A. Koot, S.M. van Ruth, A. González-Casado and L. Cuadros-Rodríguez, Proton transfer reaction-mass spectrometry volatile organic compound fingerprint for monovarietal extra virgin olive oil identification, *Food Chem.*, 134 (2012) 589–596.

[5] L.G. Dias, A. Fernandes, A.C.A. Veloso, A.A.S.C. Machado, J.A. Pereira and A.M. Peres, Single-cultivar extra virgin olive oil classification using a potentiometric electronic tongue, *Food Chem.*, 160 (2014) 321–329.

[6] C. Weihs, U. Ligges, K. Luebke, N. Raabe, *klAR Analyzing German Business Cycles*. In Baier, D., Decker, R. and Schmidt-Thieme, L. (eds.). *Data Analysis and Decision Support*, 335-343, Springer-Verlag, Berlin, 2005.

[7] J. Cadima, J.O. Cerdeira, M. Minhoto, Computational aspects of algorithms for variable selection in the context of principal components. *Comput. Stat. Data Anal.*, 47 (2004) 225-236.

[8] J. Cadima, J.O. Cerdeira, P.D. Silva, M. Minhoto, The subselect R package, (<http://cran.r-project.org/web/packages/subselect/vignettes/subselect.pdf>).

[9] W.N. Venables, B.D. Ripley, *Modern Applied Statistics with S*, Fourth Ed., Springer, New York, 2002

[10] J.M. Graham, Problems with stepwise procedures in discriminant analysis. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA, 2001.

[11] D. De March, Statistical models for high dimensional data: selection and assessment, PhD thesis, Facoltà di Economia, Università degli Studi di Firenze, 2011.