IECHo 2025 Conference

The 2nd International Electronic Conference on Horticulturae

27–29 May 2025 | Online

MDPI

Leaf spectral reflectance and machine learning for classifying olive tree cultivars in Northeastern Portugal

Pedro Marques [1,2], Luís Pádua [2,3,4], Joaquim J. Sousa [4,5] and Anabela Fernandes-Silva [1,2,3]

[1] Agronomy Department, School of Agrarian and Veterinary Sciences, University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
[2] Centre for the Research and Technology of Agro-Environmental and Biological Sciences, University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
[3] Institute for Innovation, Capacity Building and Sustainability of Agri-Food Production, University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
[4] Engineering Department, School of Science and Technology, University of Trás-os-Montes e Alto Douro, 5000-801 Vila Real, Portugal
[5] Centre for Robotics in Industry and Intelligent Systems (CRIIS), INESC Technology and Science (INESC-TEC), 4200-465 Porto, Portugal

INTRODUCTION & AIM

The accurate identification of olive tree cultivars is critical for ensuring the certification of olive oil origin and quality, conserving biodiversity, and promoting agricultural sustainability. This is of particular importance in Trás-os-Montes e Alto Douro, a region that significantly contributes to Portugal's position as the seventh-largest global producer of olive oil. Prominent cultivars in this region, such as "Cobrançosa", "Madural", and "Verdeal de Trás-os-Montes", are not only economically vital but also integral to the region's cultural heritage and traditional agricultural systems. However, traditional cultivar identification methods, including morphological characterization and molecular analysis, present notable limitations. These techniques are often time-consuming, labour-intensive, and require specialized expertise, making them unsuitable for large-scale or rapid applications.

To address these challenges, this study proposes the development of a non-destructive, efficient, and scalable methodology for olive cultivar classification. By using leaf spectral reflectance data and applying machine learning models, this research aims to improve cultivar discrimination accuracy and contribute to the advancement of precision agriculture practices.

METHOD

A total of 432 leaf samples were collected from three olive tree cultivars: "Cobrançosa", "Madural", and "Verdeal de Trás-os-Montes". Spectral reflectance measurements were performed using a spectroradiometer, covering the spectral range from 500 to 900 nm.

To address the high dimensionality of the spectral data, Principal Component Analysis (PCA) was used as a preprocessing step. This technique enabled the extraction of the most relevant spectral features while minimizing noise and redundancy, resulting in a dataset reduced to 50 principal components. For cultivar classification, four machine learning algorithms were assessed: eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Support Vector Classifier (SVC), and Decision Tree (DT). Cross-validation techniques were also applied to improve the robustness, reliability, and generalizability of the classification results.

RESULTS & DISCUSSION

The results illustrated in Table 1 revealed significant differences among the machine learning models. The XGBoost classifier showed higher performance, achieving an overall classification accuracy of 93% and outperforming the other models across all evaluation metrics. RF and SVC algorithms also showed robust classification capabilities, while the DT approach showed comparatively lower performance with an accuracy of 80%. The F1-scores also revealed distinct patterns among cultivars. The "Madural" cultivar achieved the highest F1-score of 0.95, showing its distinctive spectral signature. "Verdeal de Trás-os-Montes" followed with an F1-score of 0.93, while "Cobrançosa" showed slightly lower discriminability with a score of 0.91. This reduced performance for "Cobrançosa" can be attributed to greater spectral overlap with the other cultivars, resulting in a higher incidence of misclassifications with these two cultivars, as illustrated in Figure 1.

The results demonstrate that while machine learning models can effectively differentiate olive cultivars based on spectral data, limitations persist. The observed intra-class variability and inter-class spectral similarity, particularly between "Cobrançosa" and other cultivars, suggest that classification accuracy could potentially be improved through the integration of complementary data sources.

Model	Cultivar	Precision	Recall	F1-Score	Global accuracy
DT	Verdeal	0,82	0,72	0,77	0,80
	Madural	0,89	0,91	0,90	
	Cobrançosa	0,71	0,77	0,74	
RF	Verdeal	0,89	0,93	0,91	0,92
	Madural	0,93	0,95	0,94	
	Cobrançosa	0,93	0,86	0,89	
XGBoost	Verdeal	**0,89**	**0,95**	**0,93**	**0,93**
	Madural	**0,95**	**0,95**	**0,95**	
	Cobrançosa	**0,95**	**0,89**	**0,91**	
SVC	Verdeal	0,83	0,81	0,82	0,86
	Madural	0,91	0,93	0,92	
	Cobrançosa	0,84	0,84	0,84	

Table 1. Evaluation metrics of the machine learning models

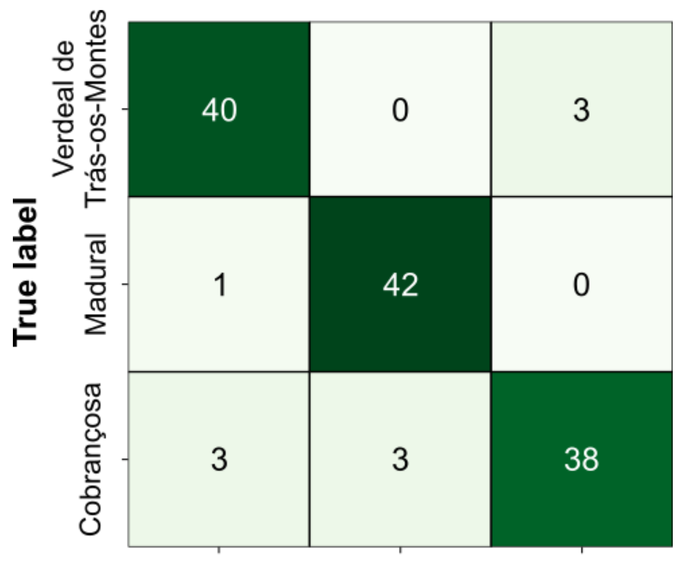

Figure 1. Confusion matrix of the XGBoost classification

CONCLUSION

This study showed the effectiveness of combining spectral reflectance with machine learning for an accurate and non-invasive identification of olive tree cultivars. The XGBoost classifier achieved the best performance with an overall accuracy of 93%, being effective for varietal classification. The approach supports more sustainable and informed decision-making in olive production sector. Future research should aim to expand the dataset to include more cultivars and varied growing conditions, integrate seasonal and multi-temporal data, and explore deep learning and hybrid models.