Optimizing Picual Olive Variety Recognition through Deep Learning and Hyperspectral Imaging in Precision Agriculture

Alba Gómez Liébana D, Ruth M. Córdoba Ortega

Researcher at the University of Jaén, Paraje Las Lagunillas Jaén, Spain e-mail: aglieban@ujaen.es | rcortega@ujaen.es

Juan J. Cubillas

Dept. Information and Communication Technologies applied to Education. International University of La Rioja, Logroño, Spain e-mail: juanjose.cubillas@unir.net

Lidia M. Ortega Dept. Computer Science. University of Jaén Jaén, Spain e-mail: lidia@ujaen.es

Abstract—The automated classification of olive varieties plays a crucial role in Precision Agriculture, enabling optimized resource allocation, improved irrigation strategies, and enhanced olive oil quality. This study explores the integration of Hyperspectral Imaging (HSI) and Deep Learning (DL) to classify olive varieties, focusing on Picual. Utilizing drone-acquired hyperspectral data, a Convolutional Neural Network (CNN) was employed to analyze leaf reflectance and extract spectral-spatial features with high accuracy. The Unmanned Aerial Vehicle (UAV)-based HSI system captures high-resolution spectral data, allowing for the detection of subtle differences in reflectance patterns that are imperceptible to traditional sensors. The study demonstrates that the proposed deep learning approach achieves an accuracy of approximately 90% in classifying olive varieties, significantly outperforming traditional machine learning methods. These findings highlight the potential of hyperspectral deep learning in agricultural applications, paving the way for scalable, efficient, and sustainable orchard management.

Keywords-Hyperspectral Imaging (HSI); Deep Learning (DL); Convolutional Neural Networks (CNN); Precision Agriculture; Olive Variety Classification; UAV-based Imaging; Spectral-Spatial Analysis; Arbequina; Picual.

I. INTRODUCTION

Olive cultivation (Olea europaea) is a fundamental component of Mediterranean agriculture, contributing significantly to global olive oil production. The identification of olive varieties is crucial for optimizing agricultural management, ensuring efficient irrigation, and enhancing oil quality. However, traditional classification methods rely on manual expertise, which is labor-intensive and impractical for large-scale olive groves [1].

Spain, with the province of Jaén as its production heart, leads the olive grove sector worldwide, being the largest producer of olive oil and a benchmark for the quality and tradition of this crop. In this province, *Picual* and *Arbequina* varieties are the most prevalent, and it is a common practice to substitute *Picual* trees with *Arbequina* due to the significant problem of Verticillium wilt. This substitution results in a high prevalence of mixed-variety groves, significantly affecting agricultural management. Specifically, irrigation, pruning, fertilization, and pest control strategies vary based on the type of variety. From a cooperative perspective, cultivar identification is crucial; *Picual* oil is characterized by an intense profile, high polyphenol content, and a bitter, pungent flavor. The identification of this variety is important to control the mixture with other varieties, such as *Arbequina*. This justifies the projects that accurately identify different tree specimens within groves.

This automatic species identification is now possible. *HSI* has emerged as a powerful tool in *Precision Agriculture*, enabling the detailed spectral analysis of plant species. Unlike multispectral imaging, *HSI* captures narrow and continuous spectral bands, allowing for the detection of subtle differences in reflectance properties between varieties [2]. This technology has been widely applied in tasks, such as vegetation monitoring, disease detection, and yield estimation [2][3]. However, conventional analysis techniques often struggle with the high-dimensional nature of hyperspectral data.

To address these challenges, Artificial Intelligence (AI) and *Deep Learning (DL)* methods have been increasingly integrated with *HSI* for agricultural applications. Deep learning techniques, particularly *CNNs*, have demonstrated significant improvements in classification accuracy for various crops, including wheat, rice, and maize [4]. *CNNs* effectively extract spectral-spatial features from hyperspectral data, reducing the need for manual feature engineering and improving classification efficiency [5].

Recent studies have highlighted the advantages of *DL* over traditional machine learning approaches in handling complex hyperspectral datasets [2]. Traditional models, such as k-Nearest Neighbors (k-NN) and Support Vector Machines (SVMs), often struggle with the curse of dimensionality and require extensive preprocessing. In contrast, *CNNs* automatically learn hierarchical feature representations, enabling superior performance in hyperspectral classification tasks [6].

Despite these advancements, limited research has been conducted on the application of deep learning for *Olive Variety Classification*. The spectral differences between olive varieties, such as *Arbequina* and *Picual*, are often subtle, making traditional classification approaches less effective [3]. Leveraging *UAV-based* hyperspectral imaging combined with

CNNs offers a promising solution for automating and scaling olive variety identification [4].

This study aims to develop a deep learning-based approach for *Olive Variety Classification* using drone-acquired hyperspectral imagery. By applying *CNN* architectures optimized for hyperspectral data, this research seeks to improve classification accuracy and provide a scalable solution for *Precision Agriculture*.

Section 2 provides an overview of related work and describes the methodology used, including data acquisition and preprocessing. Section 3 reports the experimental results, focusing on model training and performance evaluation. Section 4 discusses and interprets the findings. Finally, Section 5 presents the conclusions and outlines potential directions for future research.

II. RELATED WORK | METHODS

The research took place in Mengíbar, Jaén, on land owned by the Andalusian Institute of Agricultural and Fisheries Research and Training (IFAPA) at the Venta del Llano Center. This agricultural research facility operates under the research instituteian Regional Government and is dedicated to research and development in the agricultural sector, with a focus on olive cultivation [7]. The center is located in Jaén, which provides convenient access to a variety of olive plantations for conducting field studies and experiments. The study was carried out on a plot of land that offers optimal conditions for examining different olive varieties in a real-world agricultural setting.

The research area consists of rows of olive trees planted specifically for experimental purposes, allowing for the assessment of various olive cultivars. The experimental design includes 14 rows, each with around 24 trees. Within each row, groups of four trees from the same variety are planted, followed by a shift to a different variety. The row selected for the study can be seen in Figure 1, seeing that there are 8 *Picual* olive trees and the rest of other varieties.



Figure 1. Row selected for the study.

Most of the other varieties are hybrids under investigation and are not widely cultivated. The random arrangement of these varieties within each block ensures comprehensive data collection and reduces bias. This structure enables IFAPA to gather important insights into the adaptability, productivity, and characteristics of different olive cultivars in the specific environmental conditions of Jaén.

A. Hyperspectral data capture and preparation

This study utilized a *UAV* equipped with a NanoHyperspec camera and Light Detection and Ranging (LiDAR) sensor to acquire hyperspectral imagery of olive trees. Flight parameters

were optimized for high-quality data capture, including a 30meter altitude, 5 m/s speed, and specific overlap percentages to ensure comprehensive coverage. The hyperspectral data, capturing 270 spectral bands from 400 to 1000 nm, was processed using Headwall SpectralView[™] software, involving reflectance calibration and geometric correction using a highresolution DEM (Digital Elevation Model) generated from LiDAR data. This process resulted in a dataset of 24 olive trees, showcasing spectral variations after applying necessary corrections.

Subsequent steps focused on refining the hyperspectral data for accurate classification. Tree canopy segmentation was performed using the Enhanced Vegetation Index (EVI) and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) clustering to delineate individual trees, creating a vector mask with unique identifiers and variety classifications. To further improve data quality, noisy pixels, particularly those affected by shadows, were removed through a filtering process based on Near Infrared (NIR) reflectance and standard deviation. This filtering ensured that only spectrally stable pixels were used for analysis, enhancing the consistency and reliability of the data for subsequent classification methods. The effectiveness of these filtering techniques was demonstrated through visual comparisons and spectral signature analyses, ultimately leading to a refined dataset suitable for precise olive tree characterization.

B. Train, test and validation subsets

This section outlines the methodology for creating the training, testing, and validation subsets for **Picual vs. non-Picual (PI - NO PI)** classification. The data comes from the IFAPA farm, where the fourth row was selected for data extraction.

After segmentation, all the *Picual* olive trees were selected, totalling 264. Similarly, another 264 of the other varieties were randomly selected so that both sets were balanced, as can be clearly seen in Table I. For the training set, approximately 80% of the above-mentioned set (180 olive trees for *picual* and 186 for Non-Picual) were chosen. On the other hand, for the validation set, the remaining 20% were chosen (49 and 43 respectively), reserving 35 of each class for a subsequent test of the model generated (see Table I).

 TABLE I

 Description and distribution of Picual dataset.

Dataset	Train Data	Validation Data	Test Data	Total
PI	180	49	35	264
NO PI	186	43	35	264

This partitioning ensures that the models are trained on diverse and representative samples, improving reliability, generalizability, and reducing bias, while the separate validation set helps prevent overfitting.

C. Justification for the Use of Deep Learning and Neural Networks

Deep learning models, such as Multi-Layer Perceptrons (MLP) and *CNNs*, are particularly well-suited for handling the intricate nature of hyperspectral data. These models excel in identifying and learning hierarchical patterns directly from the data, allowing them to adapt to the complex relationships found in the numerous spectral bands of hyperspectral images. This capability is crucial when classifying olive varieties, as it enables the model to discover subtle, non-linear distinctions that might otherwise be overlooked.

Additionally, deep learning models offer the advantage of automated feature extraction, simplifying the overall process by eliminating the need for manual intervention in selecting key features. This not only streamlines the workflow but also ensures that the model can capture essential information more effectively. Combined with their ability to manage highdimensional data, deep learning models are well-equipped to improve classification accuracy and address the challenges posed by the intricate structure of hyperspectral datasets.

D. CNN architecture

As mentioned above, CNNs are highly suited for this study due to the nature of hyperspectral data and the complexity of Olive Variety Classification. A one-dimensional (1D) CNN model was developed specifically for the classification of *Picual* variety, using hyperspectral data. The model architecture consists of four convolutional layers, two max-pooling layers, a fully connected layer with dropout, and an output layer for binary classification. The structure of the model is illustrated in Figure 2. The input consists of the dataset depicted in Table I. The first convolutional layer (in dark blue) applies the Exponential Linear Unit (ELU) activation function, which improves learning and normalizes feature maps by introducing non-linearity [8]. These layers, shown in dark blue, vary in the number of filters, and their kernels scan the hyperspectral sequence to extract relevant features. Filters help capture important patterns from the data, while the kernel size remains consistent across all convolutional layers.

Following the convolutional layers, a MaxPooling layer (shown in light blue) reduces the dimensionality of the feature maps, using a pool size of 3. This step enhances the model's efficiency by focusing on the most prominent features, reducing computational complexity, and preventing overfitting. After every two convolutional layers, MaxPooling layers further reduce the dimensionality, allowing the model to retain key spectral features critical for classification.

Once the convolutional layers have extracted the necessary features, the feature maps are flattened into a one-dimensional vector and passed to a fully connected layer. This dense layer captures complex relationships between the features, applying the ReLU activation function to enhance learning by setting negative values to zero, which helps avoid issues like vanishing gradients. A dropout layer is then added to mitigate overfitting, and the final output layer uses a sigmoid function to classify the sample as either *Picual* or *Non-Picual*.

E. Computational Environment

The calculations in this study were carried out using the Anaconda distribution with Python 3.9, together with the NumPy, Pandas, TensorFlow and Scikit-learn libraries. For Bayesian optimisation, the BayesianOptimization library was used. All calculations were run on a personal computer with the following specifications: Intel(R) Core(TM) i9-12900K 12th generation 3.20 GHz processor and 64 GB of RAM. The operating system used was 64-bit on an x64-based architecture.

III. RESULTS

By employing *UAV*-based hyperspectral imaging, this study removes the necessity for manual sampling, allowing for realtime, high-throughput classification. This represents a major leap forward in *Precision Agriculture*, enhancing the scalability and efficiency of identifying olive varieties.

In this section, the hyperparameters of the *CNN* model used for classifying olive varieties are further optimized. A combination of manual tuning and Bayesian optimization was utilized to determine the most effective configurations for the *Picual* variety classification.

A. Refining the Classification

This section details the process of adjusting the hyperparameters of the *CNN* for the classification of olive tree varieties. Different configurations were explored to optimise the performance of the model, resulting in specific parameters for *Picual* variety.

Initially, hyperparameters were manually tested to improve model performance, but this approach proved to be timeconsuming and inefficient. As a result, Bayesian optimization was chosen to streamline the process and systematically explore the hyperparameter space. Bayesian optimization employs a probabilistic surrogate model to approximate the objective function—in this case, classification accuracy. It iteratively refines its search by leveraging information from previous trials, making it particularly useful when computational resources are limited or evaluations are costly, as was the case in this study.

The optimized values, detailed in Table II, include the filters applied to the convolutional layers, the kernel size for the ELU layers, and the number of neurons in the dense layer. These parameters were fine-tuned through Bayesian optimization to enhance model performance.

 TABLE II

 Range of CNN hyperparameters used for optimization in this study.

Hyperparameters	Range	
Filter 1	[10,20]	
Filter 2	[25,35]	
Filter 3	[60,75]	
Filter 4	[110,130]	
Kernel size	[2,5]	
Dense Neurons	[32,70]	

After applying Bayesian optimisation, the optimal values for the hyperparameters are shown in Table III. In addition, we



Figure 2. Neural Network CNN.

add the Pool Size, Epochs and Patience which were adjusted manually.

 TABLE III

 Range of CNN hyperparameters used for Picual.

Picual				
Hyperparameters	Value			
Filter 1	14			
Filter 2	29			
Filter 3	60			
Filter 4	111			
Kernel size	4			
Dense Neurons	59			
Pool Size	3			
Batch Size	32			
Epochs	50			
Patience	10			

B. Output of CNN

The application of *CNNs* for the classification of olive tree varieties has produced significant results, showcasing the capability of deep learning techniques to efficiently process hyperspectral data. By leveraging *UAV*-based hyperspectral imaging, this study eliminates the need for manual sampling, enabling real-time, high-throughput classification. This represents a major advancement in *Precision Agriculture*, making the identification of olive varieties more scalable and efficient. The models were evaluated based on their performance in classifying *Picual* (PI) and non-Picual (NO PI) varieties.

In addition to metrics, such as **accuracy**, **recall**, and **F1-score**, confusion matrices were generated to visualize the model's performance for each class, illustrating true positives, false positives, true negatives, and false negatives. During training, epoch plots were generated, showing the reduction in the loss function and the increase in accuracy over time, allowing for an assessment of model convergence and the detection of potential overfitting.

The *CNN* model for the *Picual* variety demonstrated robust performance:

- Loss: 0.4201
- Accuracy: 0.8804

Table IV presents the classification report for the *Picual* variety. The precision of the model shows that, when it predicts an olive tree as *Picual*, it is correct 84% of the time, while predictions of Non-Picual olive trees are correct 94% of the time. The recall metric reveals that the model accurately identifies 96% of all actual Picual olive trees and 79% of the Non-Picual ones.

The F1-Score, which provides a balance between precision and recall, is 0.90 for the *Picual* class and 0.86 for the Non-Picual class. Overall, the model achieved an accuracy of 88% in classifying the 92 test olive trees. The macro average of the metrics (precision, recall, and F1-score) represents an unweighted average across all classes, while the weighted average accounts for the number of samples per class, ensuring a more representative performance evaluation.

 TABLE IV

 Classification Report for Picual Variety.

Class	Precision	Recall	F1-Score
Picual	0.84	0.96	0.90
No Picual	0.94	0.79	0.86
Accuracy	0.88		
Macro Avg	0.89	0.87	0.88
Weighted Avg	0.89	0.88	0.88

The results of this can also be seen in the graph in Figure 3.

The epoch chart in Figure 4 visualizes how the *CNN* model improves its performance during training for the classification of the *Picual* variety. In this graph, the horizontal axis represents the training epochs, while the vertical axis shows the loss and accuracy. The confusion matrix for the *Picual* variety shown in Figure 5 presents the performance of the *CNN* model in distinguishing between *Picual* and Non-Picual (NON-PI) olive trees.

The generalisability of the model was assessed using a new set of 70 trees, equally divided between *Picual* and Non-Picual varieties. The results are shown in Table V and demonstrate the model's ability to maintain a high level of accuracy on unseen data.



Figure 3. Comparison of Recall, Accuracy, and F-score between Picual and Non-Picual.



Figure 4. Model Accuracy Across Epochs.

IV. DISCUSSION AND EVALUATION

After presenting the results obtained from the neural network for the datasets, this section provides an interpretation of those outcomes. The overall performance of the *CNN* model applied to the *Picual* variety yields promising results in terms of classification, as detailed in Table IV. The model's loss value of 0.4201 is relatively low, indicating that the network makes few errors on average when classifying the *Picual* variety. Although this loss value is slightly higher than might be ideal, it still reflects the model's effective learning.

In terms of accuracy, the model achieves 88.04%, meaning it correctly classifies *Picual* trees in the majority of cases. This level of accuracy is a solid indication of the model's capability, correctly identifying *Picual* trees 88% of the time. Regarding recall for *Picual* (see Table IV), the model demonstrates an impressive 96%, meaning it successfully identifies 96% of



Figure 5. Confusion Matrix CNN.

 TABLE V

 Classification Report for New Picual Data.

Class	Precision	Recall	F1-Score
Picual	0.78	0.91	0.84
No Picual	0.90	0.74	0.81
Accuracy	0.83 (70 instances)		
Macro Avg	0.84	0.83	0.83
Weighted Avg	0.84	0.83	0.83

all *Picual* trees in the dataset. This high recall indicates the model's strong sensitivity to the *Picual* variety, minimizing the number of *Picual* trees it misses. In contrast, the recall for Non-Picual trees is 79%, suggesting the model correctly identifies 79% of Non-Picual trees. Although lower, this value is still reasonable for distinguishing between these classes.

The F1 score, which balances precision and recall, reaches 0.90 for *Picual* and 0.86 for Non-Picual, as shown in Table IV. These high values confirm the model's strong performance across both classes, with slightly better performance for the *Picual* class. The epoch chart in Figure 4 illustrates the evolution of the model's performance during training. The gradual decrease in loss and the increase in accuracy with each epoch reflect the model's improvement over time. The curves stabilize towards the end of the training process, suggesting that the model has converged and is ready to generalize to new data.

The confusion matrix (Figure 5) offers further insight into the model's classification performance. Higher values along the diagonal indicate the model's success in correctly classifying most of the samples, while the lower off-diagonal values point to fewer misclassifications, reflecting a high level of classification reliability. To further validate the model, a new dataset of 70 trees, equally split between Picual and Non-Picual classes, was used. The model (see Table V) achieved an overall accuracy of 83% in this validation set, with a classspecific accuracy of 78% for *Picual* and 90% for Non-Picual. These results confirm the model's ability to generalize, although with slightly lower performance compared to the test set. The recall for *Picual* in the validation is 91%, while for Non-Picual it is 74%, indicating that the model is more adept at identifying *Picual* trees than Non-Picual ones. The F1 scores are 0.84 for *Picual* and 0.81 for non-Picual, demonstrating robust performance, particularly for the *Picual* variety.

V. CONCLUSION AND FUTURE WORK

This study confirms that UAV-based HSI, combined with DL, represents a highly effective solution for automated Olive Variety Classification. The CNN-based approach demonstrated strong performance in classifying the *Picual* variety with high accuracy and reliability. Nevertheless, further refinements could improve the model's robustness, including expanding the dataset to incorporate additional olive varieties and exploring how this UAV-based system adapts under various environmental conditions, such as changes in lighting and seasons. Addressing these factors would enhance the scalability and real-world application of this *Precision Agriculture* system. Unlike traditional multispectral methods, HSI enables precise differentiation of cultivars based on subtle spectral reflectance variations, significantly reducing the reliance on labor-intensive manual sampling. These findings reinforce the potential of AI-driven remote sensing for improving efficiency in *Precision* Agriculture.

The process addressed for the processing of the hyperspectral imagery includes reflectance calibration, geometric correction, and individual tree segmentation, using techniques, such as the Enhanced Vegetation Index (EVI) and the DBSCAN clustering algorithm. In addition, spectral filtering was applied to remove pixels with low reflectance, reducing noise from shaded areas in the canopy and improving the accuracy of the analysis. Then, the use of 1D *CNN* proved to be suitable for processing spectral data, with an architecture consisting of convolutional layers, max-pooling, and a fully connected layer, allowing the automatic extraction of relevant features from the data. Optimization of the *CNN* hyperparameters was crucial to obtain accurate results, with Bayesian optimization being used for the *Picual* variety.

For the *Picual* variety, the *CNN* model showed solid performance with an accuracy of 88.04% and a loss of 0.4201 in the test set, also with good generalization to unseen data. Further validation on a fresh dataset showed a slightly lower performance with an accuracy of 83%. The findings confirm that deep learning models, particularly *CNNs*, excel in extracting hierarchical spectral features from hyperspectral data, achieving significantly higher accuracy than traditional machine learning methods. Approaches, such as k-NN, Naïve Bayes, and Decision Trees struggle to handle the high-dimensional nature of hyperspectral imaging, reinforcing the superiority of data-driven feature extraction techniques in agricultural classification tasks.

Overall, the study concludes that the combination of hyperspectral imaging with deep learning is an effective tool for automated olive variety identification, which can improve agricultural practices and increase the competitiveness of olive products.

The experiments were conducted at only a single farm, so the robustness of the method should be checked on other farms as well. This limitation highlights the need to validate the proposed approach across different locations to ensure its general applicability. Also, it is suitable to verify whether the approach can be applied over longer periods and under various environmental conditions. Future research will focus on addressing the challenge of model generalization in diverse environmental conditions and crop varieties, ensuring robust performance in diverse agricultural landscapes. The evaluation of alternative *CNN* architectures, including 2D and 3D models tailored to specific data structures, will be explored. Expanding the scope to include a broader spectrum of olive varieties and integrating complementary sensor data, such as LiDAR, will improve classification accuracy and comprehensiveness. The ultimate goal would be to automatically catalog the majority species in a region using a single *UAV* flight. These advancements collectively aim to refine *Precision Agriculture* practices, promoting sustainable and efficient crop management.

ACKNOWLEDGMENTS

This research has been partially funded through the research support provided by the Ministry of Innovation and Science of the Government of Spain through the research project PID2021-126339OB-I00.

REFERENCES

- [1] E. Sena-Moreno, M. Álvarez-Ortí, D. C. Zied, A. Pardo-Giménez, and J. E. Pardo, "Olive oils from Campos de Hellin (Spain) exhibit significant varietal differences in fatty acid composition, sterol fraction, and oxidative stability", *European Journal of Lipid Science and Technology*, vol. 117, pp. 967–975, 2015. DOI: 10.1002/EJLT.201400136.
- [2] K. E. Karfi, S. E. Fkihi, L. E. Mansouri, and O. Naggar, "Classification of Hyperspectral Remote Sensing Images for Crop Type Identification: State of the Art", *Proceedings of the* 2nd International Conference on Advanced Technologies for Humanity, 2020. DOI: 10.5220/0010426600110018.
- [3] P. Messina, *Side-looking Airborne Radar (SLAR) System Operations*, Publication Title: Paula Messina, 2025.
- [4] M. Govender, K. Chetty, V. Naiken, and H. Bulcock, "A comparison of satellite hyperspectral and multispectral remote sensing imagery for improved classification and mapping of vegetation", *Water sa*, vol. 34, no. 2, pp. 147–154, 2008. DOI: 10.4314/wsa.v34i2.183634.
- [5] L. Shuai, Z. Li, Z. Chen, D. Luo, and J. Mu, "A research review on deep learning combined with hyperspectral Imaging in multiscale agricultural sensing", *Computers and Electronics in Agriculture*, vol. 217, p. 108 577, Feb. 2024, ISSN: 0168-1699. DOI: 10.1016/j.compag.2023.108577.
- [6] P. Marques, L. Pádua, J. J. Sousa, and A. A. Fernandes-Silva, "Advancements in Remote Sensing Imagery Applications for Precision Management in Olive Growing: A Systematic Review", *Remote. Sens.*, vol. 16, p. 1324, 2024. DOI: 10.3390/rs16081324.
- [7] R. G. o. A. Andalusian Research Institute, IFAPA Center "VENTA DEL LLANO" | Institute for Agricultural and Fisheries Research and Training (Instituto de Investigación y Formación Agraria y Pesquera), 2025.
- [8] Z. Khan *et al.*, "Optimizing precision agriculture: A realtime detection approach for grape vineyard unhealthy leaves using deep learning improved YOLOv7 with feature extraction capabilities", *Computers and Electronics in Agriculture*, vol. 231, p. 109 969, Apr. 2025, ISSN: 0168-1699. DOI: 10.1016/j.compag. 2025.109969.

Courtesy of IARIA Board and IARIA Press. Original source: ThinkMind Digital Library https://www.thinkmind.org