Prediction of Olive Productivity Using Machine Learning Associated with Aerial Multispectral and Thermal Imagery

Miguel Antônio Santos Faria[1], Gabriel Koch[2], Pedro Henrique Abreu Moura[2], Fábio Oseias dos Reis Silva[1], Charles Cardoso Santana[1]

[1] Minas Gerais Agricultural Research Agency – Pitangui Institute of Agricultural Technology, Pitangui, MG, 35650-000, Brazil

[2] Minas Gerais Agricultural Research Agency, Maria da Fé, MG, 37517-000, Brazil

Corresponding author: Miguel Antônio Santos Faria (miguel.antonio.itap@gmail.com)

A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Abstract.

Accurate estimation of productivity in olive orchards is fundamental for agricultural planning, resource optimization, and timely decision-making. Conventional methods for assessing productivity are labor-intensive and limited in their ability to represent spatial variability at field scale. Remote sensing using unmanned aerial vehicles (UAVs) combined with machine learning techniques offers a scalable and non-destructive approach for predicting productivity at field scale. This study evaluated the potential of spectral indices derived from multispectral and thermal imagery for predicting productivity (kg per plant) in a commercial olive orchard. Aerial surveys were conducted using a Matrice 350 RTK platform equipped with a MicaSense Altum-PT multispectral and thermal sensor. Flights were performed in February 2025, September 2025, and February 2026 at an altitude of 60 m above ground level, with 85% forward and side overlap to ensure radiometric and geometric consistency. Radiometrically calibrated orthomosaics were generated through photogrammetric processing, and 56 spectral indices associated with canopy vigor, chlorophyll content, structural properties, and physiological responses to plant stress were computed. Productivity data were obtained through field measurements. Each tree was evaluated for total production in kilograms per plant (kg/plant) during the harvest period. This continuous variable was used as the response variable for supervised machine learning regression models. Data preprocessing included variable selection and hyperparameter optimization. An optimal subset of variables was identified, including RedEdge, NIR, and thermal bands, as well as vegetation indices such as CCCI, EVI-2, ExG, SI, TGI, VVI, CI, TCARI, and MTCI. The development and evaluation of machine learning models were conducted through the following

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2026). Title of paper. In Proceedings of the 17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture and Brazilian Congress on Precision and Digital Agriculture.

stages: (1) exploratory data analysis; (2) selection of the most significant variables through feature selection; (3) hyperparameter optimization; (4) model training; and (5) evaluation of model performance using RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and R^2 metrics on an independent test dataset. Ensemble-based regression models were evaluated, including Random Forest, Gradient Boosting, XGBoost, and LightGBM. Gradient Boosting achieved the best overall performance, with RMSE of 8.88 kg/plant, MAE of 7.17 kg/plant, and R^2 of 0.72. XGBoost showed comparable performance, with RMSE of 10.36 kg/plant, MAE of 8.51 kg/plant, and R^2 of 0.67, while LightGBM obtained RMSE of 12.44 kg/plant, MAE of 10.39 kg/plant, and R^2 of 0.60. The Random Forest model showed inferior performance compared to the other ensemble models evaluated. Variable importance analysis indicated stronger contributions from the thermal band at different flight periods, followed by vegetation indices such as HUE, NIR, ExG, and EVI-2, suggesting that thermal and spectral information related to canopy vigor and plant stress are determinants in predicting productivity. The results demonstrate that integrating multispectral and thermal imagery based on UAVs with machine learning algorithms provides a viable approach for spatial prediction of productivity in olive orchards. The methodology supports precision management and may assist resource optimization strategies and early decision-making in olive production systems. Therefore, the application of machine learning with multispectral and thermal imagery for productivity prediction may represent a promising alternative for monitoring and agricultural planning in commercial olive orchards.

Keywords.
Vegetation indices, UAVs, Predictive modeling, Image processing, Remote sensing, Crop monitoring

Introduction

Olive (*Olea europaea* L.) cultivation in Brazil is a relatively recent and rapidly expanding activity, concentrated primarily in the Serra da Mantiqueira region of Minas Gerais and in the southern state of Rio Grande do Sul. Brazilian oliviculture has its scientific origins in research conducted by EPAMIG (Minas Gerais Agricultural Research Agency), which identified the Serra da Mantiqueira region at subtropical latitudes of 22–23°S and altitudes between 900 and 1,900 m — as offering the thermal and climatic conditions required for adequate chilling accumulation and reproductive development of olive trees (Coutinho *et al.*, 2017). The cultivated area has grown substantially, reaching approximately 7,000 hectares nationally, with around 2,000 hectares concentrated in the Serra da Mantiqueira alone, distributed across farms that remain predominantly small-to-medium scale and often located on steep terrain that limits mechanization (Santos *et al.*, 2017).

Despite the productive potential of the region, Brazilian olive orchards are characterized by significant interannual variability in production, driven by the alternate bearing phenomenon, climatic instability, and the still-limited availability of technical information adapted to subtropical conditions. Accurate and timely estimation of productivity at the individual plant level is therefore essential for harvest planning, resource allocation, and orchard management. However, conventional assessment methods rely on manual counting and sampling, which are time-consuming, labor-intensive, subjective, and unable to represent the spatial variability inherent to commercial orchards (Catania *et al.*, 2023).

Remote sensing using UAVs equipped with multispectral and thermal sensors offers a scalable, non-destructive, and spatially explicit alternative for monitoring orchard conditions. UAV-derived vegetation indices (VIs) have been used to characterize canopy vigor, chlorophyll content, water status, and structural properties in olive orchards, demonstrating strong associations with productive performance at the individual plant level (Catania *et al.*, 2023). When combined with ensemble machine learning (ML) algorithms such as Gradient Boosting, XGBoost, and LightGBM, these spectral and thermal features enable the development of robust predictive models for crop productivity (Bao *et al.*, 2024). The present study aimed to evaluate multispectral and thermal UAV imagery combined with ensemble ML algorithms for predicting productivity in a commercial olive orchard under subtropical conditions in Minas Gerais, Brazil.

Materials and Methods

The study was conducted in a commercial olive orchard located in the state of Minas Gerais, Brazil. Three flight campaigns were performed at key phenological dates: February 2025, September 2025, and February 2026, using a DJI Matrice 350 RTK UAV equipped with a MicaSense Altum-PT multispectral and thermal sensor, which captures six spectral bands: Blue, Green, Red, RedEdge, NIR, and Thermal Infrared (TIR). All flights were conducted at 60 m above ground level with 85% forward and side overlap, ensuring radiometric and geometric consistency across acquisition dates. Radiometric calibration was performed using a reflectance panel immediately before each flight. Calibrated orthomosaics were subsequently generated through photogrammetric processing, and mean spectral values were extracted per individual tree canopy.

A total of 56 spectral and thermal indices were computed per tree and per flight date, encompassing variables associated with canopy vigor, chlorophyll content, structural properties, and physiological stress responses. The computed variables included individual spectral bands (RedEdge, NIR, and land surface temperature - LST) and vegetation indices such as CCCI, EVI-2, ExG, SI, TGI, VVI, CI, TCARI, MTCI, and HUE, among others. Productivity data were obtained through direct field measurements, with each tree individually evaluated for total production in kilograms per plant (kg/plant) during the February 2026 harvest campaign. This continuous variable served as the response variable for all supervised ML regression models.

The modeling workflow was structured in five sequential stages: (1) exploratory data analysis to characterize variable distributions and identify potential interactions; (2) feature selection via recursive importance analysis to identify the most informative predictors; (3) Bayesian hyperparameter optimization; (4) model training on the selected variable subset with optimized configurations; and (5) evaluation of model performance on an independent test dataset using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2). Four ensemble regression algorithms were trained and compared: Random Forest (RF), Gradient Boosting (GB), XGBoost, and LightGBM.

Results and Discussion

All four ensemble models demonstrated capacity to predict olive productivity from UAV-derived spectral and thermal features. The ability of these models to capture the complex, nonlinear relationships between multi-temporal spectral and thermal variables and individual tree productivity reflects the advantage of ensemble learning approaches over simpler regression frameworks in high-dimensional agricultural datasets. The combination of three flight dates covering distinct phenological stages (post-harvest, pre-flowering, and harvest period) provided a temporally rich feature space that enabled the models to account for cumulative physiological and environmental effects on olive fruiting. The use of 56 spectral and thermal indices per tree, derived from six sensor bands including the thermal infrared channel, further expanded the predictive information available to the models. Feature selection via recursive importance analysis was essential to reduce dimensionality and retain only the most informative variables, mitigating the risk of overfitting and improving model generalization. Table 1 summarizes the performance metrics for each model on the independent test dataset.

Table 1. Performance of ensemble machine learning models for olive productivity prediction.

Model	RMSE (kg/plant)	MAE (kg/plant)	R^2
Gradient Boosting	8.88	7.17	0.72
XGBoost	10.36	8.51	0.67
LightGBM	12.44	10.39	0.60
Random Forest	-	-	<0.60

Gradient Boosting achieved the highest predictive accuracy (R^2 = 0.72; RMSE = 8.88 kg/plant; MAE = 7.17 kg/plant), followed by XGBoost (R^2 = 0.67; RMSE = 10.36 kg/plant; MAE = 8.51 kg/plant) and LightGBM (R^2 = 0.60; RMSE = 12.44 kg/plant; MAE = 10.39 kg/plant). Random

Forest showed inferior performance relative to the other ensemble models evaluated. The superiority of Gradient Boosting is consistent with previous studies reporting strong performance of this algorithm in yield prediction tasks using UAV-derived multispectral data across diverse crops (Bao *et al.*, 2024). The sequential error-correction mechanism inherent to gradient boosting is particularly suited to handling the heterogeneous and spatially variable nature of orchard productivity data.

Variable importance analysis revealed that the thermal band, across different flight periods, was the most influential predictor, followed by vegetation indices HUE, NIR, ExG, and EVI-2. The prominence of the thermal band underscores the role of canopy temperature as an integrative indicator of plant water status, photosynthetic activity, and physiological stress, all of which are closely linked to fruit set and production in olive trees. The multi-temporal approach, integrating post-harvest (February 2025), pre-flowering (September 2025), and harvest-period (February 2026) imagery, was fundamental to capturing cumulative stress signals and the dynamic spectral responses of the olive canopy throughout the productive cycle. These findings are consistent with those of Catania et al. (2023), who demonstrated strong associations between UAV multispectral imagery and olive tree production at the individual plant level, and support the use of thermal sensing as a complementary tool for olive orchard monitoring. The integration of thermal information alongside multispectral bands substantially enhanced predictive power compared to approaches relying solely on vegetation indices derived from visible and near-infrared reflectance.

Conclusion

The integration of multispectral and thermal UAV imagery with ensemble machine learning algorithms demonstrated strong potential for spatial prediction of olive productivity at the individual plant level. Gradient Boosting achieved the best performance ($R^2 = 0.72$), with thermal and spectral variables associated with canopy vigor and water stress identified as the most relevant predictors. The multi-temporal flight strategy proved essential for capturing the dynamic spectral behavior of the olive canopy throughout the production cycle. The proposed methodology supports precision management decisions, resource optimization, and harvest planning in commercial olive orchards.

Acknowledgments

The authors gratefully acknowledge the technical and institutional support of the Instituto Tecnológico de Agropecuária de Pitangui (EPAMIG ITAP) and the financial support of the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) in the development of this work.

References

Bao, Y. *et al.* (2024) 'Forecasting spring maize yield using vegetation indices and crop phenology metrics from UAV observations', *Food and Energy Security*, 13(1), p. e505. Available at: https://doi.org/10.1002/fes3.505.

Catania, P. *et al.* (2023) 'Evaluation of multispectral data acquired from UAV platform in olive orchard', *Horticulturae*, 9(2), p. 133. Available at: https://doi.org/10.3390/horticulturae9020133.

Coutinho, E.F. *et al.* (2017) 'Olive cultivation in the southern hemisphere: flowering, water requirements and oil quality responses to new crop environments', *Frontiers in Plant Science*, 8, p. 1830. Available at: https://doi.org/10.3389/fpls.2017.01830.

Santos, C.M.G. *et al.* (2017) 'Nutritional quality of olives and olive oil produced in the Serra da Mantiqueira from Brazil', *Food and Nutrition Sciences*, 8(7), pp. 735–746. Available at: https://doi.org/10.4236/fns.2017.87052.