IDENTIFICATION OF APHIDS USING MACHINE LEARNING CLASSIFIERS ON UAV-BASED MULTISPECTRAL DATA

Nathalie Guimarães (1,2), Luís Pádua (1,2,3), Joaquim J. Sousa (1,4), Albino Bento (5), Pedro Couto (1,2,3)

[1]University of Trás-os-Montes e Alto Douro (UTAD), Vila Real, Portugal
[2]Centre for the Research and Technology of Agro-Environmental and Biological Sciences (CITAB), UTAD, Vila Real, Portugal
[3]Institute for Innovation, Capacity Building and Sustainability of Agri-Food Production, UTAD, Vila Real, Portugal
[4]Centre for Robotics in Industry and Intelligent Systems, INESC-TEC, Porto, Portugal
[5]Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Bragança, Portugal

ABSTRACT

Almond trees in Portugal are susceptible to aphid infestation, which can result in reduced fruit production. To effectively tackle this issue, the combination of remote sensing (RS) data and machine learning (ML) classifiers can be used to accurately detect the presence of aphids. This study focuses in the implementation of ML classifiers and RS data analysis to identify aphids on almond trees, using high-resolution multispectral data collected through an unmanned aerial vehicle (UAV) in a Portuguese almond orchard. Four ML classifiers, kNN, SVM, RF and XGBoost, were employed and fine-tuned using vegetation indices derived from spectral data. The results revealed that the SVM classifier achieved an overall accuracy (OA) of 77%, followed by kNN with an OA of 74%, while XGBoost and RF achieved OAs of 71% and 69%, respectively. Consequently, this study demonstrates the viability of employing RS data and ML classifiers for aphid identification in almond orchards.

***Index Terms*—** Almond orchard, Vegetation indices, machine learning, support vector machine

1. INTRODUCTION

In Portugal, almonds are an important crop with considerable commercial value due to their nutritional properties. In the northeastern region of the country, almond cultivation has maintained its cultural traditions and economic importance. Almond trees (*Prunus dulcis* Mill. D. A. Webb) are vulnerable to infestation by three aphid species: Myzus persicae, Brachycaudus amygdalinus, and Brachycaudus helichrysi [1]. These insects inflict significant damage by feeding on the tree's sap, leading to reduced fruit production and, in severe cases, even tree mortality. Aphids can also transmit diseases to the tree. To prevent infestations, it is important to provide adequate soil moisture and fertilization to control weeds that can provide shelter and food for the insects. Regular tree inspections and manual removal of small aphid populations are recommended practices. In more severe cases, chemical control methods such as insecticide application may be necessary. To effectively control aphid infestations and ensure the vitality and productivity of almond trees, it is important to develop automatic detection methods that can assist farmers in predicting and managing the use of insecticides [2], [3]. Remote sensing (RS) data and machine learning (ML) techniques offer promising solutions by accurately and efficiently identify the presence of aphids, thus contributing to pest management strategies. Several studies have employed RS data, particularly vegetation indices (VIs), for the identification and classification of pests and diseases in various crops. For instance, Kumar et al. [4] found a significant correlation between aphid infestation in mustard crops and specific spectral indices such as the Normalized Difference Vegetation Index (NDVI), Red-Green Vegetation Index (RVI), and Chlorophyll Absorption Index (CAI). Elliott et al [5], identified an effective waveband and spectral VI derived from two different wavelengths for detecting aphid-damaged wheat. Riedell and Blackmer [6] found that leaf reflectance in certain areas, along with the Normalized Total Pigment to Chlorophyll Ratio Index (NPCI), proved to be reliable indicators of chlorophyll loss in wheat caused by aphid feeding. Conversely, Mirik et al [7] observed that the Aphid Index (AI) consistently exhibited a significant correlation with Russian wheat aphid abundance in all fields, while other indices, such as the NDVI, SIPI, and DSSI, were less effective. To the best of our knowledge, no studies have employed RS data for aphid detection on almond trees. To address this research gap, we propose a study aimed at developing a method to identify aphids on almond trees using ML classifiers and RS data.

2. MATERIAL AND METHODS

2.1. Study area and UAV data collection

The study was caried out on July 4, 2022, in a rain-fed almond orchard encompassing 111 trees. Situated in São Salvador, Mirandela, in the Trás-os-Montes region of northeastern Portugal (41°25'55" N, 7°08'06' W, as shown in Fig. 1), the orchard region has a dry, hot climate suitable for almond growth.

Fig. 1. Overview of the studied almond orchard and its location within Portugal mainland.

Data acquisition was performed using a Phantom 4 Pro V2.0 UAV (DJI, Shenzhen, China), equipped with an equipped with a 20-megapixel (MP) CMOS sensor. The sensor was mounted on a three-axis gimbal to ensure stable imaging and georeferencing of the RGB images. Multispectral images were acquired using the Sequoia multispectral (Parrot SA, Paris, France). This sensor enabled the acquisition of images across four distinct bands with a resolution of 1.2 MP: green (550 nm - 40 nm bandwidth), red (660 nm - 40 nm bandwidth), red edge (735 nm - 10 nm bandwidth) and near Infrared (790 nm - 40 nm bandwidth). The radiometric calibration of the acquired images was performed using irradiance data collected during flight with a solar irradiance sensor, along with images from a specific calibration target. UAV flights were conducted at a height of 60 m, with 80% longitudinal imagery overlap and 70% lateral overlap to ensure comprehensive coverage and produce a dense 3D point cloud.

2.2. Data Processing

The data processing was divided into four main steps (Fig. 2). Step 1 encompassed the photogrammetric processing of RGB and multispectral data, resulting in the creation of a digital surface model (DSM), a digital terrain model (DTM), a tree canopy height model (CHM) (calculated by subtracting the DTM from the DSM), an orthophotomosaic, and VIs. Step 2 focused on tree canopy delineation, which involved creating a tree canopy mask. This was achieved through a segmentation process that applied a height threshold to the CHM. Step 3 involved feature extraction and labeling, i.e., extracting the VIs and spectral bands associated with each tree canopy object. Furthermore, a labeling process was conducted to classify the objects of each tree canopy. Finally, in step 4, ML classifiers were trained and tested using the labeled dataset.

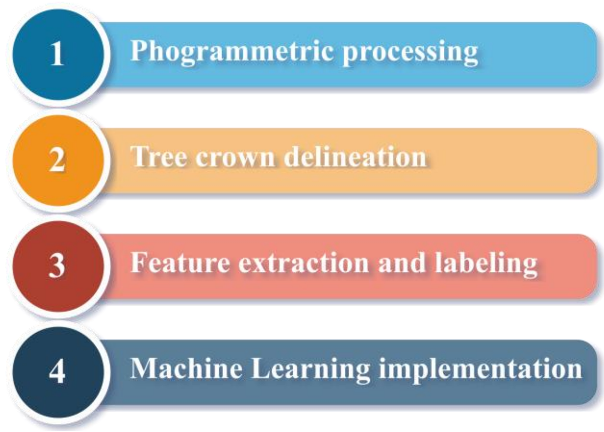

Fig. 2. Data processing main steps: (1) Photogrammetric processing; (2) Tree crown delineation; (3) Features extraction and labeling; (4) Machine Learning implementation.

2.2.1. Photogrammetric processing

The acquired UAV images, containing both RGB and multispectral data, were processed using photogrammetric techniques implemented in Pix4Dmapper Pro (Pix4D SA, Lausanne, Switzerland). This software applies radiometric calibration to ensure accurate multispectral data analysis and uses the Structure from Motion (SfM) techniques to reconstruct the 3D structure of the surveyed scene from a series of images, resulting in the generation of a 3D point cloud from the acquired 2D images. The alignment of the images was achieved by detecting common points and incorporating ground control points (GCPs). The outputs of this process included various raster products, such as the DSM, DTM, orthophoto mosaic (derived from the RGB images), and VIs (computed from the multispectral images). Subsequently, the DSM and DTM were processed in QGIS software to calculate the canopy height model (CHM).

2.2.2. Tree crown delineation

Accurate delineation of tree crowns is essential for obtaining tree-level parameters. In this study, tree crown delineation was performed by segmenting individual canopies based on the CHM. A threshold of 0.5 m was applied for binarization, resulting in a binary mask that was vectorized to create polygons representing each detected tree crown.

2.2.3. Feature extraction and labeling

During the feature extraction and labeling step, different VIs and spectral bands associated with each tree crown were extracted using the Raster Statistics for Polygons tool provided by System for Automated Geoscientific Analyses (SAGA). Mean values of VIs and spectral bands were calculated for each object, considering the information of each pixel within it. This process generated a dataset with multiple features. For the application of ML classifiers, a target variable (class field) was created on the dataset, providing information about aphid infestation. Within this process, two labels were established for the class field, distinguishing trees with aphids detected (class with label 1) from trees without detected aphids (class with label 2).

2.2.4. Machine Learning implementation

The ML implementation involved several steps, including data collection and preparation, classifier selection (training and testing), performance evaluation, and hyperparameter tuning. A dataset was created, consisting of 16 features associated with VIs and spectral bands (ExNIR, ExRE, CIRE, EVI2, GN, GNDVI, GREEN, GRVI, NDRE, NDVI, NIR, OSAVI, REDEDGE, REN, RN, and SAVI). The dataset contained a target variable "class", which included the labels for detected and not detected aphid infestation. The dataset was analyzed to detect missing values, assess feature distribution, and detect outliers. Features normalization was performed using StandardScaler from the Scikit-learn Python library. Pearson's correlation coefficient was also implemented to remove irrelevant or redundant data, resulting in the selection of four features (GN, GREEN, GRVI, and NDRE). Subsequently, four classifiers, namely kNN, SVM, RF and XGBoost, were implemented. The classifiers were trained (with 75% of the dataset samples) and tested (with 25% of the dataset samples) with the selected features, and their performance was evaluated using several metrics, including precision, recall, f1-score and overall accuracy (OA).

3. RESULTS AND DISCUSSION

As shown in Fig. 3, GN, GREEN, GRVI, and NDRE emerged as important features for classification. Notably, NDRE and GRVI exhibit a more distinct separability, thereby potentially enhancing the accuracy of the classification process.

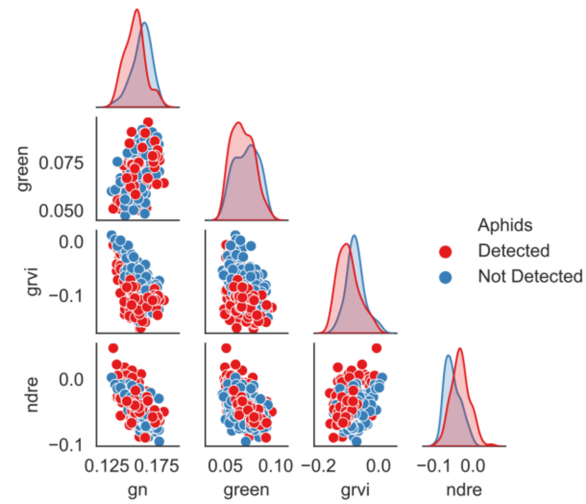

Fig. 3. Distribution of the target variable among the selected features. The diagonal plots show the Kernel Density Estimator (KDE) for each feature.

Based on the results obtained by the various classifiers evaluated (Table 1), the SVM presents the best performance with an OA of 77%, followed by kNN with an OA of 74%, while XGBoost demonstrated an OA of 71%. In comparison, the RF classifier showed the poorest performance, achieving an OA of 69%. It is noteworthy that all classifiers performed well, surpassing the baseline OA of 51% obtained from a simple dummy classifier that does not learn from the data [8].

Table 1. Performance evaluation of kNN, SVM, RF and XGBoost in the classification of aphid infestation. Aphids detected are represented by class 1 and aphids not detected are represented by class 2.

Classifiers	Classes	Precision	Recall	F1-Score	OA
kNN	1	0.69	0.83	0.75	0.74
	2	0.81	0.67	0.73	
SVM	1	0.73	0.83	0.77	0.77
	2	0.83	0.73	0.77	
RF	1	0.62	0.90	0.73	0.69
	2	0.85	0.52	0.64	
XGBoost	1	0.63	0.93	0.75	0.71
	2	0.89	0.52	0.65	

4. CONCLUSIONS

The findings of this study demonstrate the potential of RS and ML techniques in pest detection and control in agriculture. In particular, aphids pose a substantial threat to almond orchards, inflicting damage to trees and reducing crop productivity. Nevertheless, traditional methods employed for aphid detection and monitoring tend to be time-consuming and often unreliable. The use of RS data, such as high-resolution UAV-based aerial imagery, in conjunction with ML classifiers, enabled the accurate identification of aphid infestations in almond orchards.

This technology has the potential to significantly improve cropping practices by providing near real-time monitoring of pest populations, allowing for more precise and targeted application of pest control measures. This can lead to the mitigation of higher losses, reduced treatment costs, and a more sustainable approach to farming.

ACKNOWLEDGMENTS

Financial support was provided by national funds through FCT – Portuguese Foundation for Science and Technology (UI/BD/150727/2020), under the Doctoral Programme "Agricultural Production Chains – from fork to farm" (PD/00122/2012) and from the European Social Funds and the Regional Operational Programme Norte 2020. This study was also supported by CITAB research unit (UIDB/04033/2020), Inov4Agro (LA/P/0126/2020) and by CIMO (UIDB/00690/2020).

5. REFERENCES

[1] J. Benhadi-Marin, J. A. Pereira, J. A. Barrientos, A. Bento, e S. A. P. Santos, «Diversity of predaceous arthropods in the almond tree canopy in Northeastern Portugal: A methodological approach», *Entomological Science*, vol. 14, n.º 3, pp. 347–358, 2011, doi: 10.1111/j.1479-8298.2011.00444.x.

[8] E. van de Bijl, J. Klein, J. Pries, S. Bhulai, M. Hoogendoorn, e R. van der Mei, «The Dutch Draw: Constructing a Universal Baseline for Binary Prediction Models». arXiv, 24 de março de 2022. doi: 10.48550/arXiv.2203.13084.

[2] W. Almatni e N. Khalil, «A primary survey of aphid species on almond and peach, and natural enemies of Brachycaudus amygdalinus in As-Sweida, Southern Syria».

[3] M. Â. Rodrigues, «Manual Técnico da Amendoeira: Estado da Produção». CNCFS, 2017.

[4] J. Kumar, A. Vashisth, V. K. Sehgal, e V. K. Gupta, «Identification of Aphid Infestation in Mustard by Hyperspectral Remote Sensing», *Journal of Agricultural Physics*, vol. 10.

[5] N. Elliott *et al.*, «Airborne Remote Sensing to Detect Greenbug1 Stress to Wheat», *swen*, vol. 34, n.º 3, pp. 205–211, set. 2009, doi: 10.3958/059.034.0301.

[6] W. E. Riedell e T. M. Blackmer, «Leaf Reflectance Spectra of Cereal Aphid-Damaged Wheat», *Crop Science*, vol. 39, n.º 6, pp. 1835–1840, 1999, doi: 10.2135/cropsci1999.3961835x.

[7] M. Mirik, G. J. Michels, S. Kassymzhanova-Mirik, e N. C. Elliott, «Reflectance characteristics of Russian wheat aphid (Hemiptera: Aphididae) stress and abundance in winter wheat», *Computers and Electronics in Agriculture*, vol. 57, n.º 2, pp. 123–134, jul. 2007, doi: 10.1016/j.compag.2007.03.002.