A MACHINE LEARNING APPROACH TO DETECT DEAD TREES CAUSED BY LONGHORNED BORER IN EUCALYPTUS STANDS USING UAV IMAGERY

André Duarte[1,2], Nuno Borralho[2], Mário Caetano[1,3]

[1]NOVA Information Management School (NOVAIMS) Universidade Nova de Lisboa
Campus de Campolide, 1070-312, Lisbon, Portugal
[2]Forest and Paper Research Institute (RAIZ)
Quinta de São Francisco, 3800-783, Eixo, Aveiro, Portugal
[3]Direção Geral do Território (DGT)
Rua da Artilharia 1 107, 1099-052, Lisbon, Portugal

This is the accepted version of the conference paper published by IEEE at IGARSS 2021 . 2021 IEEE International Geoscience and remote sensing Symposium: Proceedings.

How to cite: Duarte, A., Borralho, N. and Caetano, M. (2021). A Machine Learning Approach to Detect Dead Trees Caused by Longhorned Borer in Eucalyptus Stands Using UAV Imagery," In IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium: Proceedings (pp. 5818-5821). IEEE. https://doi.org/10.1109/IGARSS47720.2021.9554947.

A MACHINE LEARNING APPROACH TO DETECT DEAD TREES CAUSED BY LONGHORNED BORER IN EUCALYPTUS STANDS USING UAV IMAGERY

André Duarte[a,b], Nuno Borralho[b], Mário Caetano[a,c]

[a]NOVA Information Management School (NOVAIMS) Universidade Nova de Lisboa
Campus de Campolide, 1070-312, Lisbon, Portugal
[b]Forest and Paper Research Institute (RAIZ)
Quinta de São Francisco, 3800-783, Eixo, Aveiro, Portugal
[c]Direção Geral do Território (DGT)
Rua da Artilharia 1 107, 1099-052, Lisbon, Portugal

ABSTRACT

Pest damages in eucalyptus plantations cause significant economic losses for the pulp and paper industry. Longhorned borers (ELB) outbreaks induce mortality in eucalyptus stands. In this study, multispectral imagery was obtained from unmanned aerial vehicles. We attempt to improve the classification process done in previous work. The local maxima of sliding a window and the Large-Scale Mean-Shift segmentation (LSMS) were applied to extract tree crows. Subsequently, the mean of spectral bands and twelve vegetation indices were calculated to characterize each segment. To classify tree canopies into dead and healthy trees, supervised machine learning using Random Forest (RF) and Support Vector Machine (SVM) were applied. The overall accuracy of Random Forests was 98.35% and Support Vector Machine of 97.7%. We concluded that SVM did not perform better than RF. Moreover, adding new vegetation indices in the classification process did not increase accuracy.

***Index Terms*—**Eucalyptus stands, Longhorned borer (ELB), unmanned aerial vehicles (UAV), machine learning.

1. INTRODUCTION

Outbreaks of insect pests such as Longhorned borers (ELB) (*Phorachanta semipunctata* and *Phoracantha recurve)* cause Eucalyptus globulus mortality across regions with a Mediterranean climate. The assessment of the damage caused by ELB is usually not detected using traditional survey techniques until significant mortality has occurred [1]. Nevertheless, some symptoms such as canopy collapse, insect exit holes or kyno bleeding might indicate ELB attacks [1,2]. The best approach to reduce the impact of ELB is early detection. Therefore, UAV remote sensing can be a promising tool for forest managers and practitioners to planning and control actions.

Combining traditional survey techniques with UAV remote sensing can minimize the response time and reduce the losses. Those combinations have become employed for pest and disease monitoring, such as bark beetle [3][4], using UAV technology and non-parametric algorithms for classification. For instance, Nasi [3] investigate the bark beetle damage at the tree level using object-based K-Nearest Neighbor (KNN) classification. Honkavaara et al. [4] explore multitemporal imagery to study spectral symptoms using a Random Forest machine learning model to classify spruce health. Duarte et al. [5] studied the mortality of ELB in *Eucalyptus globulus* stand using the mean of five spectral indices and four bands. To classify the canopies Random Forest classifier was used. However, the potential for pest monitoring has still not been fully exploited, and detection algorithms are missing for most pests.

This study aims to use twelve spectral indices and four bands to assess the impacts of ELB attacks using the same imagery collected in the previous work [5]. On the other hand, Random Forest with selected hyperparameters and Support Vector Machine algorithms are tested. Our focus is classifying the dead trees on eucalypt stands and compare the models generated.

2. METHODOLOGY

2.1. Study site

The study site is part of a *Eucalyptus globulus* stand near the village of Gavião in central Portugal (39º27.909’ N, 7º56.124’W) (Figure 1). This area is managed by The Navigator Company (NVG), a Portuguese pulp and paper company. It has an area of approximately 30.56 ha and is composed of about 28,271 trees. The mortality caused by ELB was detected in 2018. The severe drought in 2017 induced the intensity of attacks.

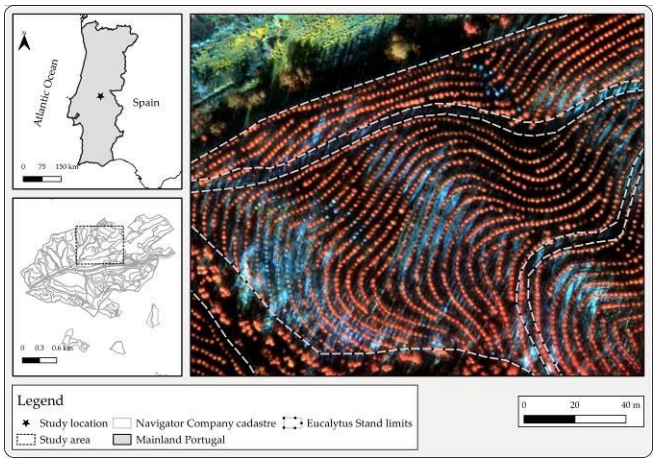

Figure 1. Location of the study site and a highlight of multispectral image.

The site has a Mediterranean climate with hot dry summers and mild winters, and the altitude varies between 160 and 250m a.sl.(above sea level).

2.2. Data acquisition

Multispectral images were collected using the Parrot Sequoia camera mounted on an eBee SenseFly (Parrot S.A., Paris, France) [6]. The camera collects four discrete bands with 1.2 megapixels resolution: green (530-570 nm), red (640-680nm), red-edge (730-740nm) and near-infrared (770-810nm). Four autonomous flights were conducted on 21 January 2019 with a height of 190 m above terrain level, speed was 10 m/s, and image overlap of 80%. The mission planning application used was SenseFly eMotion (Parrot S.A., Paris, France). For geometric correction purposes, nine ground control points (GCP) were collected using Arrow Gold Real Time Kinematics (RTK) antenna.

To generate the orthomosaic an Ag Multispectral template pre-defined on Pix4Dmapper Pro software (Version 4.2, Pix4D S.A., Prilly, Switzerland) [7] was used. The spatial resolution of multispectral imagery is 17 centimeters. Field data were acquired at the tree level, and health status assessment was assigned in two categories: dead and healthy.

2.3. Data processing

The object-based analysis was performed using the Large Scale Mean Shift algorithm (LSMS) provided by Orfeo Toolbox [8]. Subsequently, the mean of the four bands and vegetation indices for each feature was calculated. For this proposal, twelve spectral indices were derived (Table 1). The local maxima algorithm was applied to extract tree position, available in the QGIS Tree density plugin [9]. Given the imagery, segmented and tree positions were intersected to obtain the tree crowns described in the last work done [5].

Table 1 - Vegetation indices calculated.

Vegetation index	Equation	References
Difference vegetation index (DVI)	DVI = NIR-Red	[10]
Green Normalized Difference Vegetation Index (GNDVI)	GNDVI = NIR-Green/NIR+Green	[11]
Normalized Difference Red-Edge (NDRE)	NDRE = (NIR-RedEdge)/(NIR+RedEdge)	[12]
Normalized Difference Vegetation Index (NDVI)	NDVI = NIR-Red/NIR+Red	[13]
Soil Adjusted Vegetation Index (SAVI)	SAVI = 1.5 x (NIR-Red)/(NIR + Red+0.5)	[14]
Green-Red Vegetation Index (GRVI)	GRVI = Green-red/Green + red	[13]
Anthocyanin Reflectance Index (ARI)	ARI = (1/Green)-(1/RedEdge 1)	[15]
Modified Anthocyanin Reflectance Index (mARI)	mARI = (1/Green)-(1/NIR)*NIR	[15]
Red Green Index (RGI)	RGI = Red/Green	[16]
Modified Anthocyanin Content Index (mACI)	mACI = (1/Green)-(1/Rededge)*NIR	[15]
Chlorophyll Index (CI)	CVI = (NIRxRed)/$(Green)^2$	[17]
Infrared Percentage Vegetation Index (IPVI)	IPVI = NIR/(NIR+Red)	[18]

We select 2027 out of 25911 segments based on field data and on-screen interpretation to train the algorithms. To classify tree canopies into two different classes, healthy and dead trees were selected.

2.4. Machine Learning classifiers

In order to classify tree canopies into two different classes, we applied two machine learning algorithms, the Random Forests (RF) presented by Breiman [19] and Support Vector Machine (SVM) by Vapinik [20]. In the previous work done, we used default values provided by OrfeoToolbox for RF, which were maximum tree depth 5, and the number of trees was 100. For these experiments, the RF hyperparameters selected were a maximum tree depth, 100 trees, and Gini criteria (Table 2). Larger trees help to convey more info, whereas smaller tree gives less precise information. Therefore, a considerable depth enough to split each node to the desired number of observations. The hyperparameters for SVM were linear kernels, the decision function shape was one-vs-one (OVO), and the C parameter was 1. The decision function shape OVO trains a classifier for each 2-pair class combination. The C parameter defines the margin to avoid misclassifying each training example. Hence, C small values correspond to a large margin hyperplane, even if the hyperplane misclassifies more points.

The dataset of training areas was partitioned into 70% for training and 30% to test. Experimental tests were implemented in IPython environment using Scikit-Learn Python library [21].

Table 2 - Hyperparameters for each algorithm.

Classifier	Hyperparameters	Values
RF	Maximum tree depth	None
	Trees	100
	criterion	Gini
SVM	Kernel	Linear
	Decision function shape	one-vs-one
	C parameter	1

2.5. Evaluation metrics

The error matrix was performed to determine the agreement between the classified segments and the ground truth image regarding trees dead and healthy. From the error matrix, we calculated the overall accuracy (OA), producer's accuracy (PA) or recall, and user's accuracy (UA) or precision. Furthermore, the Kappa statistic was also estimated to evaluate the agreement between classification and ground truth data quantitively [22]. The supervised classification in each algorithm was validated with 2% of randomly stratified selected segments using the research tools provided by QGIS software.

3. EXPERIMENTAL RESULTS

RF and SVM reached an overall accuracy higher than 90%. Alvaréz-Taboada[23] stressed that using two classes is suitable from a management point of view. On the other hand, increasing the number of classes might be difficult to discriminate the different infestation levels. With RF classifier, the overall accuracy was 98.3% (Table 3), while for SVM was 97.7% (Table 4).

Analyzing the confusion matrix of RF user's accuracy (precision) for healthy trees was 98.9%, and the class dead trees was 93.1% (Table 3). The producer's accuracy (recall) for dead trees was 91.5% and healthy trees of 99.1%. These results showed that a healthy tree is the most distinguished class.

Table 3 - Confusion matrix of the RF classification

Class		Predicted				
		Healthy	Dead	Total	PA(%)	Kappa
Observed	Healthy	455	4	459	99.1	0.91
	Dead	5	54	59	91.5	
	Total	460	58	518	-	
	UA(%)	98.9	93.1	-	98.3	

Regarding the confusion matrix of SVM user's accuracy (precision) for healthy trees was 99.3%, and the class dead was 85.5% (Table 4). The producer's accuracy (recall) for dead trees was 94.6% and healthy trees of 98.1%. These results showed that the healthy tree was the most distinguished class using SVM.

Table 4 – Confusion matrix of SVM classification.

Class		Predicted				
		Healthy	Dead	Total	PA(%)	Kappa
Observed	Healthy	454	9	463	98.1	0.89
	Dead	3	53	56	94.6	
	Total	457	62	519	-	
	UA(%)	99.3	85.5	-	97.7	

RF and SVM overall accuracy were very similar. However, at the class level, the user accuracy of dead trees in the RF experiment (93.1%) was significantly higher than SVM (85.5%).

Figure 2a) represents the spatialization of the classification of tree canopy status. Figure 2 b) and 2 c) shows highlights of RF and SVM classifications, respectively. Figure 2 b) and 2 c) the healthy trees have red color, and dead one blue or white.

When compared to a previous experiment in the same study area [5], this research demonstrated that SVM did not perform better than RF. Furthermore, adding new vegetation indices in the classification process did not increase accuracy.

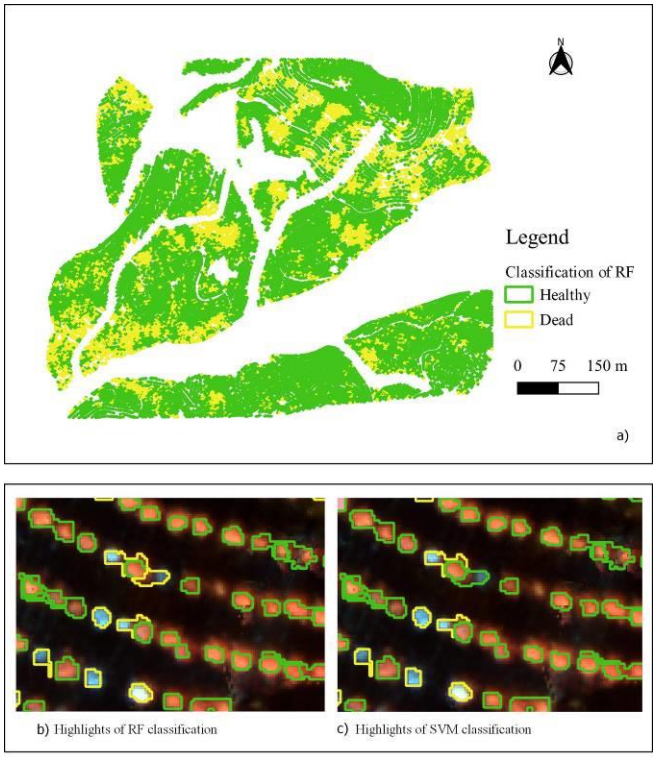

Figure 2. Classification of tree canopy status: a) a general overview of RF classification, b) highlights of RF classification, c) highlights of SVM classification.

4. CONCLUSION

Mortality in *Eucalyptus globulus* at the stand level can be successfully mapped using UAV imagery. Combining the segmentation and local maxima filter to extract the individual-tree crowns helps to discriminate the two canopies. Machine learning approaches can support forest managers and practitioners to identify stressors in their stands. The overall accuracy of RF classification was 98.3%, and SVM of 97.7%. Using other hyperparameters and adding new vegetation indices in the classification process did not increase accuracy. Future work focuses on adding new additional forested sites affected by ELB to compare different edafo-climatic conditions. Further research might focus the study at different spatial scales.

5. ACKNOWLEDGEMENTS

We would like to thank Terradrone and all colleagues of the RAIZ team. The presented work was also carried out with a research project financed by the MySustainableForest project, which has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 776045.

5. REFERENCES

[1] Wotherspoon, K.; Wardlaw, T.; Bashford, R.; Lawson, S. Relationships between annual rainfall, damage symptoms and insect borer populations in midrotation Eucalyptus nitens and Eucalyptus globulus plantations in Tasmania: Can static traps be used as an early warning device? Aust. For., 77, 15–24, 2014.

[2] Seaton, S.; Matusick, G.; Ruthrof, K.X.; Hardy, G.E.J. Outbreaks of Phoracantha semipunctata in response to severe drought in a Mediterranean Eucalyptus forest. Forests, 6, 3868–3881, 2015.

[3] Näsi, R.; Honkavaara, E.; Lyytikäinen-Saarenmaa, P.; Blomqvist, M.; Litkey, P.; Hakala, T.; Viljanen, N.; Kantola, T.; Tanhuanpää, T.; Holopainen, M. Using UAV–Based Photogrammetry and Hyperspectral Imaging for Mapping Bark Beetle Damage at Tree–Level. Remote Sens., 7, 15467 15493, 2015.

[4] Honkavaara, E., Näsi, R., Oliveira, R., Viljanen, N.,Suomalainen, J., Khoramshahi, E., Haataja, L.. Using multitemporal hyper-and multispectral UAV imaging for detecting bark beetle infestation on norway spruce. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 43(B3), 429–434, 2020.

[5] Duarte, A.; Acevedo-Muñoz, L.; Gonçalves, C.I.; Mota, L.; Sarmento, A.; Silva, M.; Fabres, S.; Borralho, N.; Valente, C. Detection of Longhorned Borer Attack and Assessment in

[6] SenseFly Parrot Group. Parrot Sequoia Multispectral Camara. Available online: https://www.sensefly.com/camera/parrotsequoia/ (accessed on january 2021).

[7] Pix4D. Pix4D—Drone Mapping Software. Version 4.2. Available online: https://pix4d.com/ (accessed on january 2021).

[8] OTB Development Team. OTB CookBook Documentation; CNES: Paris, France, p. 305, 2018.

[9] Crabbé, A.H.; Cahy, T.; Somers, B.; Verbeke, L.P.; Van Coillie, F. Tree Density Calculator Software Version 1.5.3, QGIS. 2020. Available online: https://bitbucket.org/kul-reseco/localmaxfilter (accessed on january 2021).

Eucalyptus Plantations Using UAV Imagery. *Remote Sens.*, *12*, 3153, 2020.

[10] Richardson, A.J.; Weigand, C.L. Distinguishing vegetation from background information. Photogramm. Eng. Remote Sens., 43, 1541–1552, 1977.

[11] Gitelson, A.A.; Merzlyak, M.N. Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll. J. Plant Physiol., 148, 494–500, 1996.

[12] Barnes, E.M.; Clarke, T.R.; Richards, S.E.; Colaizzi, P.D.; Haberland, J.; Kostrzewski, M.;Waller, P.; Choi, C.; Riley, E.; Thompson, T.; et al. Coincident detection of CropWater Stress, Nitrogen Status and Canopy Density Using Ground-Based Multispectral Data [CD Rom]. In Proceedings of the Fifth International Conference on Precision Agriculture, Bloomington, MN, USA, 16–19 July 2000.

[13] Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens.
Environ., 150, 127–150, 1979

[14]Huete, A.R. A soil adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309, 1988.

[15] Gitelson, A. A., Chivkunova, O. B., & Merzlyak, M. N. Nondestructive estimation of anthocyanins and chlorophylls in anthocyanic leaves. American Journal of Botany, 96, 1861–1868, 2009.

[16] Gamon , J. , and J. Surfus . Assessing leaf pigment content and activity with a refl ectometer. *New Phytologist* 143 : 105 – 117, 1999 .

[17] Vincini, M., Frazzi, E., *and* D'Alessio, P.. A broad-band leaf chlorophyll index at the canopy scale. Precis. Agric. 9:303–319, 2008.

[18] Crippen R. Calculating the vegetation index faster. Remote Sens Environ 34:71–73, 1990.

[19]Breiman, L. Random forests. Mach. Learn. 45, 5–32, 2001.

[20]Vapnik, V.; Golowich, S.E.; Smola, A.J. Advances in neural information processing systems. In Support Vector Method for Function Approximation, Regression Estimation and Signal Processing, Proceedings of the NIPS'96 9th International Conference on Neural Information Processing Systems, Denver, CO, USA, 3–5 December 1996; MIT Press: Cambridge, MA, USA, 1996.

[21] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, _E. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12 (Oct):2825-2830, 2011.

[22] Congalton, R.G.; Green, K. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices; CRC Press;
Taylor & Francis Group: Boca Raton, FL, USA, ISBN 1498776663, 2019.

[23] Álvarez-Taboada, F., Sanz-Ablanedo, E., Rodríguez Pérez, J.R., Castedo-Dorado, F., Lombardero, M.J. Multi-sensor and multi-scale system for monitoring forest health in *Pinus radiata* stands defoliated by *Lymantria dispar* in NW Spain. *Proceedings of the ForestSAT Open Conference System*, http://ocs.agr.unifi.it/index.php/forestsat2014/ForestSAT2014/paper/view/245, 2014.