Species classification of cork oak and stone pine trees using airborne laser scanning and individual tree detection

D. N. Cosenza[1], P. Soares[1], M. Tomé[1], J. M. C. Pereira[1], J. M. N. Silva[1]

[1]Forest Research Centre, School of Agriculture, University of Lisbon, Tapada da Ajuda, 1349-017 Lisbon,
Email: dncosenza@gmail.com;{paulasoares; magatome; jmcpereira; joaosilva}@isa.ulisboa.pt

1. Introduction

Cork oak (*Quercus suber* L.) woodlands are important ecosystems in Mediterranean countries. They provide wood and non-wood materials, regulate water quality, prevent soil erosion, and provide cultural services for the community (Bugalho et al. 2011). This ecosystem is particularly valuable for Iberian economies, where 80% of global cork are produced (50% in Portugal, and 30% in Spain, APCOR 2019). Such importance made Portugal implement rigid protection laws to control exploitation of the cork oaks. The felling of cork oaks might be punishable by a fine, and authorized cuts (e.g., for road construction) must be compensated by the plantation of trees in an area 1.25 times larger than the intervened area. However, cork oak trees are frequently mixed with stone pines (*Pinus pinea* L.), which are used for cone and pine kernel production. The spatial heterogeneity of both species in the stands creates difficulties to traditional forest inventory. An alternative is using remote sensing techniques to collect tree-level data. In this case, individual tree detection (ITD) using remote sensing must be used along with species classification algorithms.

Airborne laser scanning (ALS) and ITD data are widely applied for tree species classification (Fassnacht et al. 2016). The process involves isolating trees in the point clouds and computing metrics to be used as predictors in classification models. Different approaches can be used for supervised classification, namely the linear discriminant analysis (LDA), k-nearest neighbors (kNN), random forest (RF), artificial neural networks (ANN), and support vector machines (SVM) – see Korpela et al. (2010) and Deng et al. (2016). Most research compared these approaches for the case of boreal and temperate forests. However, to the best of our knowledge, there is still limited information regarding their effectiveness in Iberian woodlands. Thus, this study aims to benchmark different classification approaches to distinguish between cork oak and stone pine trees in pure and mixed stands. We tested LDA, kNN, RF, ANN, and SVM assessing for classification accuracy with different training data sizes.

2. Methods

The study area was in the Alentejo region, in mid-south Portugal. The forest stands were in powerline wayleaves (Figure 1a). High-density ALS data (>45 returns m^{-2}) were collected using a helicopter flying at low altitude. ALS data analysis was conducted using *lidR* package (Roussel et al. 2020), so please see the package documentation for further details about ALS data processing. Trees were segmented based on Silva et al. (2016) algorithm. Visual inspection was conducted to select 1000 cork oaks and 1000 stone pines trees to build the training data (Figure 1b). Packalén et al. (2012) simulated annealing algorithm was used to select 15 predictor metrics (Table 1). The selections were based on the Kappa coefficient, where Kappa=(p_o-p_e)/(1-p_e), p_o is the relative observed agreement, and p_e is the probability of chance agreement. LDA was trained using the *MASS* package (Venables and Ripley 2002), kNN with *yaImpute* (Crookston and Finley 2008), RF with *randomForest* (Liaw and Wiener 2002), ANN with *nnet* (Venables and Ripley 2002), and SVM with *e1071* (Meyer et al. 2020). kNN was trained using k=5, inverse distance weighting, and distance metric computed using Euclidean distance (kNN_Euc), Mahalanobis (kNN_Mah), most similar neighbor (kNN_MSN), and random forest (kNN_RF). ANN was trained using a single hidden layer with 10 neurons. All other model hyperparameters were set to package default.

Random and balanced samples of the original dataset were used to compare models. We tested training data sizes of 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 600, 800, 1000, 1500, and 2000 trees. The models were compared using the Kappa statistic and overall accuracy (i.e., percentage of agreement) computed by 10-fold cross-validation repeated 100 times for each training data size.

Published in: Markus Hollaus, Norbert Pfeifer (Eds.): Proceedings of the SilviLaser Conference 2021, Vienna, Austria, 28–30 September 2021.
Technische Universität Wien, 2021. DOI: 10.34726/wim.1861
This paper was peer-reviewed. DOI of this paper: 10.34726/wim.1967

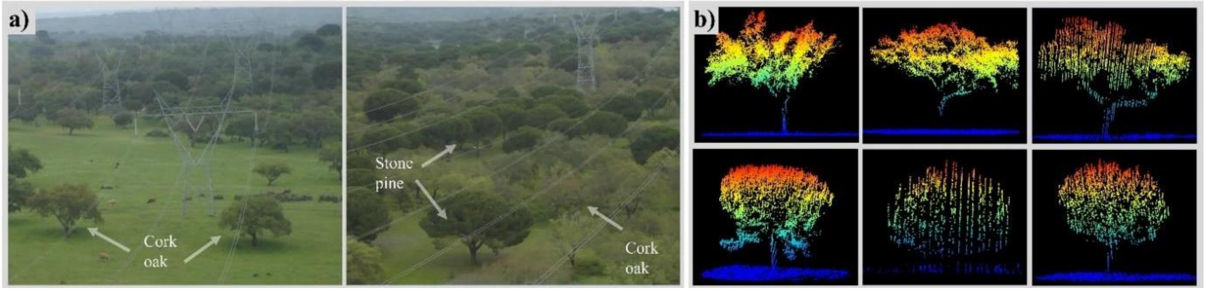

Figure 1: a) Images of the area; b) Examples of cork oak (top row) and stone pine (bottom row) trees.

Table 1. ALS metrics used in the models.

Type	Metrics[*]
LDA	h_{35}, h_{45}, h_{60}, h_{80}, h_{95}, h_{d3}, h_{d8}, h_{smean}, h_{max}, h_{sd}, h_{kurt}, i_{dq90}, i_{sd}, *curv*, *line*
kNN_Euc	h_{20}, h_{45}, h_{55}, h_{85}, h_{90}, h_{d4}, h_{d6}, h_{sd}, h_{amean}, h_{smean}, h_{cmean}, h_{skew}, i_{max}, i_{dq90}, *plan*
kNN_Mah	h_{10}, h_{45}, h_{75}, h_{85}, $h_{90,}$ h_{d1}, h_{d4}, h_{d6}, h_{d7}, h_{d8}, h_{a2m}, i_{dq30}, i_{mean}, i_{skew}, *sphe*
kNN_MSN	h_{d5}, h_{d7}, h_{d8}, h_{d9}, h_{mean}, h_{amean}, h_{cmean}, h_{sd}, h_{a2m}, i_{mean}, i_{sd}, $i_{skew,}$ λ_m, *line*, *sphe*
kNN_RF	h_5, h_{15}, h_{60}, h_{85}, h_{d2}, h_{d3}, h_{max}, h_{a2m}, i_{dq10}, i_{kurt}, i_{mean}, i_{sd}, *line*, *sphe*, *hori*
RNA	h_5, h_{10}, h_{20}, h_{40}, h_{75}, h_{d1}, h_{d2}, h_{d6}, h_{d7}, h_{d9}, h_{dq70}, i_{sd}, i_{skew}, *hori*, *plan*
RF	h_{30}, h_{40}, h_{85}, h_{90}, h_{d2}, h_{d6}, h_{d7}, h_{mean}, i_{sd}, h_{cv}, i_{dq10}, i_{dq90}, i_{kurt}, i_{sd}, *line*
SVM	h_{30}, h_{35}, h_{50}, h_{60}, h_{65}, h_{85}, h_{d8}, h_{skew}, i_{max}, i_{sd}, i_{dq10}, λ_l, *line*, *sphe*, *anis*

[*] Prefixes *h* and *i* indicate return height and intensity metrics and subscripts indicate the following statistics: x-th percentile (*x*), maximum (*max*), mean (*mean*), square mean (*smean*), cubic mean (*cmean*), kurtosis (*kurt*), skewness (*skew*), standard deviation (*sd*), coefficient of variation (*cv*), interquartile distance range (*iqr*), percentage below the x-th height fraction (*dx*) in a total of 10 fractions, percentage of returns above 2 m (*pa2m*), percentage or returns above mean (*amean*), percentage below the x-th height percentile (*qx*); vertical complexity index (*vci*); eigen-based metrics: (λ_s), medium (λ_m), and largest (λ_l) eigen values, anisotropy (*aniso*), curvature (*curv*), horizontality (*horiz*), linearity (*line*), planarity (*plan*), and sphericity (*spher*).

3. Results and discussion

Each approach had similar patterns for Kappa and overall accuracy (Figure 2). kNN_RF performed the best but comparable to RF and SVM. ANN had an intermediary performance. LDA, kNN_Mah, and kNN_MSN had poor and similar performances, while kNN_Euc performed the poorest. All approaches improved performance rapidly when more training data were used, but marginal improvements were noted after 400-600 training trees. For instance, when using >600 trees kNN_RF, RF, and SVM had Kappa values between 0.70-0.75 and 85-87% of accuracy, ANN had between 0.65-0.70 for Kappa and 82-84% of accuracy, and the others between 0.48-0.58 and 74-79%.

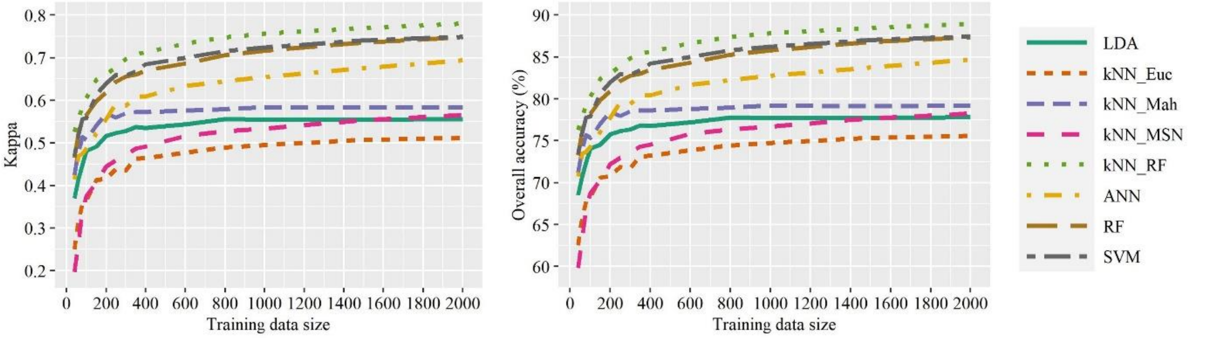

Figure 2: Kappa coefficient (left) and overall accuracy (right) for different training data sizes.

kNN_RF has been successfully used to classify tree species in boreal forests (Vauhkonen et al. 2010), and SVM, RNA, and RF in temperate forests (Deng et al. 2016). The advantage of RF and kNN_RF is their ability to handle high-dimensional datasets. However, Åkerblom et al. (2017) obtained better results with kNN_Euc distance than with SVM. Differently from Åkerblom et al. (2017), we did not tune the model hyperparameters, which might explain our poor performance for kNN_Euc. The minimal training data size was in line with the literature, where Korpela et al. (2010) found accuracies around 87% using kNN_MSN and just 300 training trees in boreal forests. Other factors might affect the accuracy of the classification algorithms, such as the point density, ALS metrics, and forest structure. Our analysis involved only two classes, so it is likely that less effective results are achieved if more species are included. The effect of the ITD algorithm was also not considered, but the experience suggests this would not be significant if well-calibrated algorithms are used. Furthermore, our study involves sparse broad leave and conifer trees, so it is possible that analyses based on satellite images also provide satisfactory results for a more cost-effective inventory. All these topics must be addressed in further studies for the case of woodlands and mixed stands of cork oak and other species.

4. Conclusion

kNN_RF, RF, and SVM were the best models to distinguish cork oak from stone pine trees using ALS and ITD. Balanced training data of 400 trees allowed training models with Kappa ≥0.7 and overall accuracy >83%.

Acknowledgments

We thank Rede Elétrica Nacional, SA (REN) for the datasets. This research was funded by the Forest Research Centre, a research unit funded by Fundação para a Ciência e a Tecnologia I.P. (FCT), Portugal (UIDB/00239/2020). D. N Cosenza was also supported by FCT [PD/BD/128489/2017].

References

Åkerblom, M., Raumonen, P., Mäkipää, R., and Kaasalainen, M. 2017. Automatic tree species recognition with quantitative structure models. *Remote Sensing of Environment* 191: 1–12. doi:10.1016/j.rse.2016.12.002.

APCOR [Portuguese Cork Association]. 2019. Information bureau: cork sector in numbers. Available from https://www.apcor.pt/wp-content/uploads/2019/02/CORK-SECTOR-IN-NUMBERS_EN.pdf [accessed 5 June 2021].

Bugalho, M.N., Caldeira, M.C., Pereira, J.S., Aronson, J., and Pausas, J.G. 2011. Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services. *Frontiers in Ecology and the Environment* 9(5): 278–286. doi:10.1890/100084.

Crookston, N.L., and Finley, A.O. 2008. yaImpute: an R package for κNN imputation. *Journal of Statistical Software* 23(10): 1–16. doi:10.18637/jss.v023.i10.

Deng, S., Katoh, M., Yu, X., Hyyppä, J., and Gao, T. 2016. Comparison of tree species classifications at the individual tree level by combining ALS data and RGB images using different algorithms. *Remote Sensing* 8(12): 1034. doi:10.3390/rs8121034.

Fassnacht, F.E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., and Ghosh, A. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment* 186: 64–87. doi:10.1016/j.rse.2016.08.013.

Korpela, I., Ørka, H., Maltamo, M., Tokola, T., and Hyyppä, J. 2010. Tree species classification using airborne LiDAR – effects of stand and tree parameters, downsizing of training set, intensity normalization, and sensor type. *Silva Fennica* 44(2): 319–339. doi:10.14214/sf.156.

Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. Available from https://cran.r-project.org/package=randomForest [accessed 14 September 2019].

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., and Leisch, F. 2020. e1071: misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. Available from https://cran.r-project.org/package=e1071 [accessed 4 March 2021].

Packalén, P., Temesgen, H., and Maltamo, M. 2012. Variable selection strategies for nearest neighbor imputation methods used in remote sensing-based forest inventory. *Canadian Journal of Remote Sensing* 38(5): 557–569. doi:10.5589/m12-046.

Roussel, J.-R., Auty, D., Coops, N.C., Tompalski, P., Goodbody, T.R.H., Meador, A.S., Bourdon, J.-F., de Boissieu, F., and Achim, A. 2020. lidR: an R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sensing of Environment* 251: 112061. Elsevier. doi:10.1016/j.rse.2020.112061.

Silva, C.A., Hudak, A.T., Vierling, L.A., Loudermilk, E.L., O'Brien, J.J., Hiers, J.K., Jack, S.B., Gonzalez-Benecke, C., Lee, H., Falkowski, M.J., and Khosravipour, A. 2016. Imputation of individual longleaf pine (*Pinus palustris* Mill.) tree attributes from field and LiDAR data. *Canadian Journal of Remote Sensing* 42(5): 554–573. doi:10.1080/07038992.2016.1196582.

Vauhkonen, J., Korpela, I., Maltamo, M., and Tokola, T. 2010. Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sensing of Environment* 114(6): 1263–1276. Elsevier Inc. doi:10.1016/j.rse.2010.01.016.

Venables, W.N., and Ripley, B.D. 2002. Modern applied statistics with S. *In* 4th edition. Springer New York, New York, USA. doi:10.1007/978-0-387-21706-2.