

EVALUATION OF XGBOOST AND LGBM PERFORMANCE IN TREE SPECIES CLASSIFICATION WITH SENTINEL-2 DATA

H. Ło's¹, G. Sousa Mendes¹, D. Cordeiro¹, N. Grosso¹, H. Costa^{2,3}, P. Benevides², M. Caetano^{2,3}

¹Deimos Engenharia, Avenida Dom João II 41 10o, Lisbon, Portugal

²Direção Geral do Território, Rua Artilharia 1 107, Lisbon, Portugal

³NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 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: Los, H., Mendes, G. S., Cordeiro, D., Grosso, N., Costa, H., Benevides, P., & Caetano, M. (2021). Evaluation of Xgboost and Lgbm Performance in Tree Species Classification with Sentinel-2 Data. In *IGARSS 2021 - 2021 IEEE International Geoscience and Remote Sensing Symposium: Proceedings* (pp. 5803-5806). IEEE. <https://doi.org/10.1109/IGARSS47720.2021.9553031>

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

EVALUATION OF XGBOOST AND LGBM PERFORMANCE IN TREE SPECIES CLASSIFICATION WITH SENTINEL-2 DATA

H. Loś¹, G. Sousa Mendes¹, D. Cordeiro¹, N. Grosso¹, H. Costa^{2,3}, P. Benevides², M. Caetano^{2,3}

¹Deimos Engenharia, Avenida Dom João II 41 10º, Lisbon, Portugal

²Direção Geral do Território, Rua Artilharia 1 107, Lisbon, Portugal

³NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, Lisbon, Portugal

ABSTRACT

Tree species classification with satellite data has become more and more popular since Sentinel-2 launch. We compared efficacy and effectiveness of Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LGBM) with widely used in remote sensing Random Forest (RF), Support Vector Machine (SVM) and K-Nearest Neighbour (KNN) algorithms. Analyses were performed over an area in Portugal with multi-temporal Sentinel-2 data registered in April, June, August and October 2018. The selected classes were: cork oak, holm oak, eucalyptus, other broadleaved, maritime pine, stone pine and other coniferous. Algorithm efficacy was measured through F1-score and accuracy while efficiency was measured through the median time needed for each fit. XGBoost and LGBM outperformed efficacy of other algorithms, which was already high (above 90% for the best variant of each algorithm). In terms of efficacy, LGBM overcame all algorithms, including XGBoost.

Index Terms— Tree species classification, NextLand, Sentinel-2, XGBoost, LGBM, Random Forest, Support Vector Machine, K-Nearest Neighbours

1. INTRODUCTION

Information about tree species is indispensable element of forest inventories [1]. Traditionally, this information is collected through sampling during field works and later extrapolated to larger areas. For over 30 years, satellite data have been tested as auxiliary data to support forest inventories [2]. Sentinel-2 data are particularly useful because they contains more near shortwave infrared channels compared to earlier high and medium spatial resolution satellites. There is rich literature on tree species classification based on Sentinel-2 data, where significant impact of data timeseries is shown [3–9]. In these publications mainly two algorithms are used,

This work is part of the project funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 869520.

Random Forest (RF) [3–8] and Support Vector Machine (SVM) [8,9], which are well known and have been used in remote sensing for many years. Often, for classification of satellite images, K-Nearest Neighbour (KNN) algorithm is used together with SVM and RF as benchmarking algorithms [10,11].

Along with the development of machine learning, new algorithms for processing of large data amounts have been constantly developed. The goal of our research was to analyze the efficacy and efficiency of algorithms Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LGBM) comparing to RF, SVM and KNN, when classifying 8 classes of tree species in Portugal based on Sentinel-2 data. Both XGBoost [12] and LGBM [13] are algorithms that, as RF, are based on an ensemble of decisions trees, but use gradient boosting to increase their performance. LGBM was proposed as a faster alternative to XGBoost.

2. DATA AND METHODS

2.1. Reference data

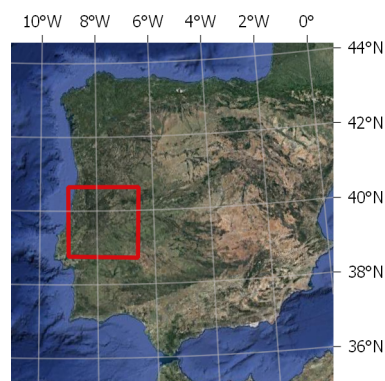


Fig. 1. Area of study: central Portugal, covered by Sentinel-2 tiles SND, SPD, TNE, TPE.

The study area covers approximately 200 km by 200 km

region in the central part of Portugal (fig. 1). We focused on 5 broadleaved classes - cork oak (*Quercus suber*), holm oak (*Quercus rotundifolia*), various type of eucalyptus (*Eucalyptus*) divided into two group: young and adult, other broadleaved - and 3 coniferous - maritime pine (*Pinus pinaster*), stone pine (*Pinus pinea*), other coniferous. These classes represent the main species and types of forest trees in Portugal.

The reference data were gathered using automatic extraction of about 6000 random points for each class from auxiliary datasets such as the national land cover and land use map (COS). Other auxiliary datasets were used to improve the land cover data filtering such as national burnt areas maps and Copernicus High-Resolution Layers (HRL) [14].

2.2. Satellite data

Sentinel-2 is a constellation of twin satellites flying in the same orbit with phased at 180° . This gives a high revisit frequency of 5 days or better, depending on latitude. Satellites provide wide-swath, high-resolution, multi-spectral images in 13 spectral bands.

We generated timeserie from four tiles (29SND, 29SPD, 29TNE, 29TPE) at level 2A, registered in April, June, August and October 2018. We selected the following Sentinel-2 bands: visible (B2, B3, B4), vegetation red edge (B5, B6, B7), near-infrared (B8, B8A) and shortwave infrared (B11, B12). Bands with 20 m pixel size were resampled to 10 m, so all bands had the same pixel size.

2.3. Methods

Algorithms were trained and tested with a pipeline (fig 2) divided in 4 stages: (i) data are shuffled and separated into train and test, preserving the initial distribution of classes. The first set is used to train and validate the algorithms, and the second to infer (through approximation) the true error. (ii) pre-processing stage, or feature scaling, where the model is fitted with training data and used to scale the test data, and saved. (iii) a random grid search [15], for each algorithm, to fit several different configuration of hyper-parameters (parameters that cannot be learnt from data and are set *a priori*). (iv) training and cross validation of each configurations. The model with the best set of hyper-parameters is then evaluated with test data.

In stage (i) we set the split at 60% for training, and 40% for test, which makes 27828 pixels and 18552 pixels, respectively. It gives approximately 3479 pixels to train each of 8 classes. In (ii) for pre-processing of each algorithm we used normalization (`_NORM`) and standardization (`_STD`), and, additionally, no scaling for tree models. In stage (iii) we set 100 random grid searches to test the following hyper-parameters:

- KNN: between 3 and 10 neighbours;
- RF: maximum depth of trees between 3 and 30, spaced by 2,

number of estimators in [100, 250, 500, 750, 1000], maximum number of features to consider at each split between 10% and 40%;

- SVM: radial basis function kernel with C parameter varying between 1 and 30;

- XGBoost: maximum depth of trees between 3 and 30, spaced by 2, number of estimators in [100, 250, 500, 750, 1000], learning rate in [0.02, 0.05, 0.1, 0.2, 0.3].

- LGBM: maximum depth of trees between 3 and 30, spaced by 2, number of estimators in [100, 250, 500, 750, 1000], learning rate in [0.1, 0.15, 0.2, 0.25, 0.3, 0.4], number of leaves in [10, 20, 30, 50, 100, 125, 150, 175, 200].

Note that for the KNN and SVM, the random grid search is an exhaustive grid search, since there are not enough parameters to make 100 combinations, only 8 and 30, respectively. Finally, for the cross validation in stage (iv), we set $k = 5$.

To increase the efficacy, we used One-Vs-All approach, in which a model is trained individually for each class. Making the problem binary greatly decreasing the complexity of each model. This results in 8 models, in total, where the final classification is given by the model with highest score.

The efficacy was measure through the F1-score, the harmonic mean between precision and recall, with the first being defined as the ratio of true positives against the sum of true positives with false positives, and the second as the ratio of true positive against the sum of true positives with false negatives. We also used accuracy, that is defined as the ratio between the sum of true positives with true negatives and total number of samples. Having both of these metrics allowed us to be more confident on the results, since they score the efficacy differently.

The efficiency of an algorithm was measure through the median time of each fit in stage (iii).

Tests were performed on Ubuntu 18.04, with an Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz, with 10 cores and 24 Gb of RAM. Code was written in Python, using the libraries provided by the community. All algorithms were parallelized in CPU mode.

3. RESULT AND DISCUSSION

3.1. Efficacy

We obtained very good results of efficacy, above 90% for F1-score and accuracy values for the best pre-processing variant, for all algorithms (fig. 3). Out of the five algorithms, the highest F1-score and accuracy value was obtained for LGBM (96% for both measures) and slightly lower for XGBoost (95% for both measures), SVM (94% for both measures), KNN (F1-score: 91%, accuracy: 92%) and RF (92% for both measures).

For all algorithms, by applying standardization in the pre-processing stage, we reached higher F1-scores and accuracy values than in the case when normalization was used. It is

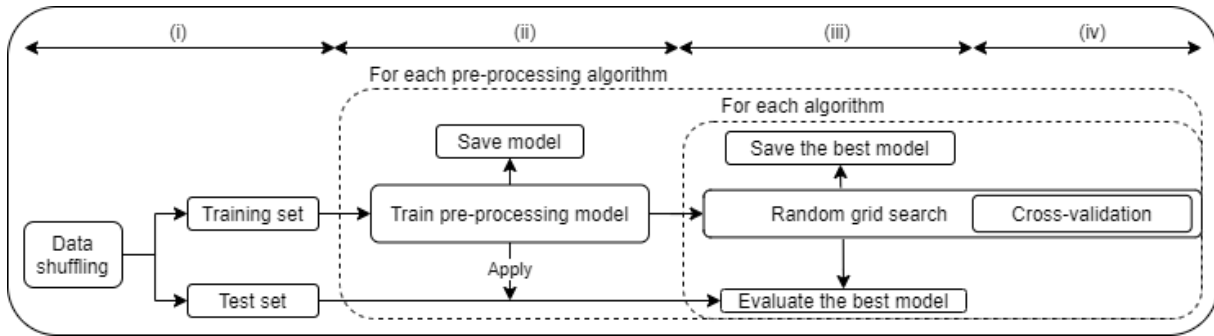


Fig. 2. Pipeline used to run and compare the algorithms divided into four stages: (i) data shuffling and split, (ii) pre-processing, (iii) random grid search, (iv) cross validation (inside a grid search).

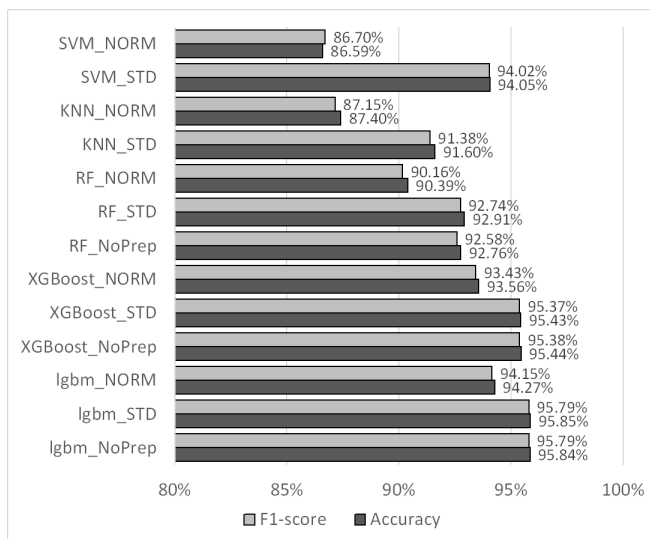


Fig. 3. Efficacy measures: F1-score and accuracy. Suffix at algorithm name refers to pre-processing.

especially visible for the SVM algorithms, with the efficacy measures higher by 7%, and to a lesser extent for KNN and RF, with the efficacy measures higher by 4% and 2%, respectively. The differences in the values of the efficacy measures due to the use of standardization in the pre-processing step or the lack of pre-processing are negligible (less than 1%).

3.2. Efficiency

Figure 4 presents the median time to fit each algorithm to the data. The LGBM algorithm turned out to be the fastest of all algorithms, with one fifth of the time needed by XGBoost and SVM. LGBM also has one of the lowest standard deviations, together with KNN, making them fairly reliable.

Even if SVM takes more time per fit than the LGBM, in total it took less time to achieve high scores, since it needed fewer iterations on our setup, as stated in Section 2.3, opposed to the LGBM that had to run the 100 iterations. However,

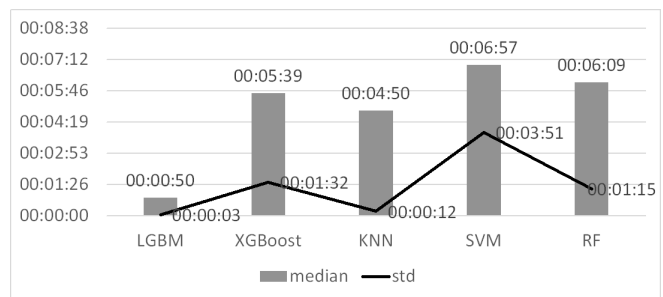


Fig. 4. Efficiency measures: median and standard deviation from fits (time in hours).

analyzing those 100 fits of LGBM, Figure 5, we can observe how it constantly is able to achieve high scores. If we randomly compare 30 iterations of LGBM (equivalent of setting 30 in stage (iii)), we still get better results than the SVM, even if we consider the last 30 iterations when ranked, where the first of those had 94,33% in F1-Score.

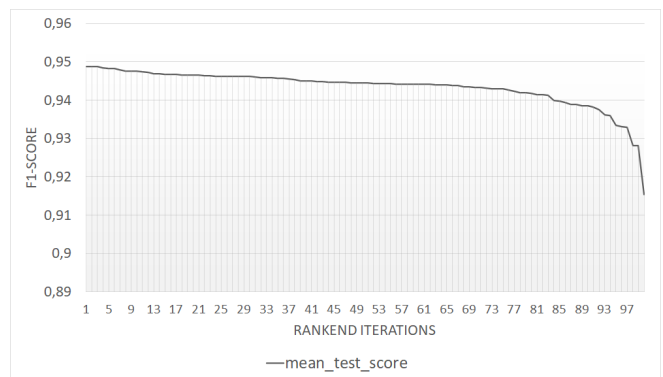


Fig. 5. Plot with the ranked F1-score of all 100 fits of LGBM_NoPrep

4. CONCLUSIONS

The aim of presented study was to analyze efficacy and efficiency of XGBoost and LGBM algorithms for tree species classification comparing to RF, SVM and KNN. The results show that LGBM and XGBoost are good alternatives to the benchmarking algorithms when analysing efficacy. When including efficiency, LGBM overcame XGBoost and the benchmarking algorithms. As mentioned in the Introduction, LGBM was developed with the intention of XGBoost improvement in terms of speed, and in our study the significant improvement has been observed. When comparing to SVM (the best of the selected benchmarking algorithms in terms of efficiency), LGBM was more stable. Moreover, SVM requires feature scaling, which is not the case of LGBM. In our work, we tested the most relevant hyper-parameters and we plan to include other hyper-parameters in further studies. As the next step, comparison of the algorithms presented in this paper will be performed with reference data presenting tree species from different climate zones. Presented work will be developed in a service under Horizont 2020 project NextLand (<https://ec-nextland.eu/>).

5. REFERENCES

- [1] A. Kangas and M. Maltamo, *Forest inventory: methodology and applications*. Springer Science & Business Media, 2006, vol. 10.
- [2] R. Fontaine, Ed., *Manual of forest inventory*. FAO. [Online]. Available: <http://www.fao.org/3/a-ap358e.pdf>
- [3] M. Persson, E. Lindberg, and H. Reese, "Tree species classification with multi-temporal sentinel-2 data," *Remote Sensing*, vol. 10, no. 11, p. 1794, 2018.
- [4] E. Grabska, P. Hostert, D. Pflugmacher, and K. Ostapowicz, "Forest stand species mapping using the sentinel-2 time series," *Remote Sensing*, vol. 11, no. 10, p. 1197, 2019.
- [5] M. Immitzer, M. Neuwirth, S. Böck, H. Brenner, F. Vuolo, and C. Atzberger, "Optimal input features for tree species classification in central europe based on multi-temporal sentinel-2 data," *Remote Sensing*, vol. 11, no. 22, p. 2599, 2019.
- [6] D. Wang, B. Wan, P. Qiu, Y. Su, Q. Guo, R. Wang, F. Sun, and X. Wu, "Evaluating the performance of sentinel-2, landsat 8 and pléiades-1 in mapping mangrove extent and species," *Remote Sensing*, vol. 10, no. 9, p. 1468, 2018.
- [7] J. Lim, K.-M. Kim, E.-H. Kim, and R. Jin, "Machine learning for tree species classification using sentinel-2 spectral information, crown texture, and environmental variables," *Remote Sensing*, vol. 12, no. 12, p. 2049, 2020.
- [8] M. Wessel, M. Brandmeier, and D. Tiede, "Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on sentinel-2 data," *Remote Sensing*, vol. 10, no. 9, p. 1419, 2018.
- [9] A. Y. Denisova, L. M. Kavelenova, E. S. Korchikov, N. V. Prokhorova, D. A. Terentyeva, and V. A. Fedoseev, "Tree species classification for clarification of forest inventory data using sentinel-2 images," in *Seventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2019)*, vol. 11174. International Society for Optics and Photonics, 2019, p. 1117408.
- [10] L. A. Vega Isuhuaylas, Y. Hirata, L. C. Ventura Santos, and N. Serrudo Torobeo, "Natural forest mapping in the andes (peru): A comparison of the performance of machine-learning algorithms," *Remote Sensing*, vol. 10, no. 5, p. 782, 2018.
- [11] P. Thanh Noi and M. Kappas, "Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery," *Sensors*, vol. 18, no. 1, p. 18, 2018.
- [12] G. Ke, Q. Meng, T. Finely, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems 30 (NIP 2017)*, December 2017. [Online]. Available: <https://www.microsoft.com/en-us/research/publication/lightgbm-a-highly-efficient-gradient-boosting-decision-tree/>
- [13] T. Chen and C. Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, aug 2016.
- [14] I. Hernandez, P. Benevides, H. Costa, and M. Caetano, "Exploring sentinel-2 for land cover and crop mapping in portugal," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 83–89, 2020.
- [15] J. Bergstra and Y. Bengio, "Random search for hyperparameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. null, p. 281–305, Feb. 2012.