Random Forest Algorithm and Convolutional Neural Networks for the Tree Species Classification in Remote Sensing Data

Zdeněk Patočka[1], Petr Strejček[1], Anton Malyshev[1]

[1]Department of Forest Management and Applied Geoinformatics, Faculty of Forestry and Wood Technology, Mendel University in Brno, Zemědělská 1165/1, 613 00 Brno-Černá Pole, Czech Republic

Abstract

This study deals with the classification of tree species using modern methods of machine- and deep-learning applied to satellite and drone date The aim of the study is to demonstrate the ability of these methods to accurately identify and classify different tree species. The first part is focused on the use of DeepForest and Detectree2 algorithms for tree crown delineation, which allow efficient segmentation and detection of trees in complex aerial images. The work with YOLO (You Only Look Once) algorithm is presented, the purpose of which is to train a model for specific detection and classification of selected tree species from drone data. The results of this algorithm is compared to the results of Random Forest machine learning algorithm. Second part of the study is focused on tree species classification in the large area of the University Forest Area by using of the Sentinel-2 and PlanetScope data. It was used the Random Forest algorithm and permanent sample plot to train the algorithm a to create map of the main tree species.

Keywords: *deep learning, machine learning, multispectral data, drone, satellite, PlanetScope, Sentinel-2, Detectree2, YOLO*

1 Introduction

Currently, forestry management and ecological research are increasingly turning to technologies such as unmanned aerial systems (drones), remote sensing data, and machine learning methods to more effectively map and monitor forest ecosystems. Tree species maps can then serve as supplementary outputs for forest management plans and as baseline data for clustering lidar data, enabling the application of regression models specific to the main tree species.

https://doi.org/10.11118/978-80-7701-024-5-0047

2 Material and Methods

In the first part of the study, data was acquired using the senseFly eBee Plus drone and the Parrot Sequoia multispectral camera and the S.O.D.A photogrammetric camera. The data were processed in the Agisoft Metashape software into the form of a multispectral orthophoto, vegetation indices and digital surface models. At the same time, ground truth data was collected using the Trimble R12i GNSS receiver. Object annotation was then performed using CVAT.

DeepForest and Detectree2 algorithms were used to identify crowns. The number of trees and the shape of their crowns were then compared with the inverse watershed segmentation method. Subsequently, a convolutional neural network, specifically the YOLO algorithm, was used to search for individual tree species. The results were compared with the Random Forest method.

For the creation of the main forest types, the Sentinel-2 and PlanetScope data in vegetation and non-vegetation period has been downloaded. The representation of tree species weighted by basal area was calculated for each sample plot. Sample plots containing at least 80% of the given tree species were selected as train samples. Band values and vegetation indices were calculated for these areas.

Subsequently, the Random Forest algorithm was trained in R and deployed on the entire territory of the UFE Křtiny. This created a map of the main tree species.

3 Results

The accuracy of tree crown mapping was determined by comparing the area of individual crowns with a standard that was created based on the manual marking of crowns. In case of use algorithm DeepForest could not perform this comparison because this algorithm it does not create precise polygons surrounding the tree crowns and thus does not allow for precise

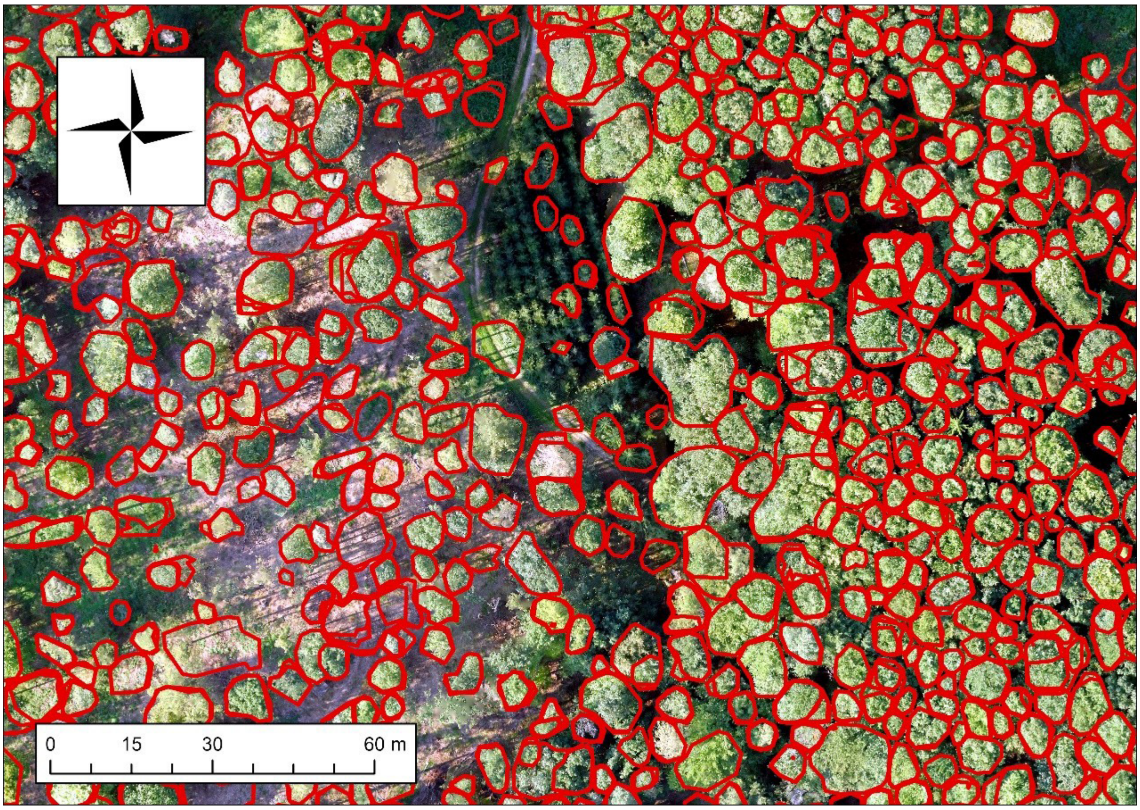

Fig. 1: *Identified tree crowns using Detectree2 algorithm*

Category	F1 – score	Recall	Precision	Producer‘s Accuracy	User‘s Accuracy
Oak	0,76	0,71	0,83	0,892	0,961
Spruce	0,76	0,62	1,0	0,952	0,833
Pine	0,6	0,6	0,6	0,9	0,818
Beech	0,62	0,71	0,55	0,822	0,948
Larch	0,82	1,0	0,7	0,923	0,827

Tab. 1 *Evaluation of the classification performed by Random Forest and IWS*
Kappa Index: 0,854
Overall Accuracy: 0,885

determination their area. In contrast, the Detectree2 algorithm achieved a match of 96% between automatically calculated crown area and manual marking (with a tolerance of ±30% in area crowns). The IWS method showed an accuracy of 40%, which means that only 67 crowns of the total number had a comparable content to manually marked crowns.

Two crown delineations (IWS and Detectree2) were performed for use in the Random Forest method, as it was assumed that the quality of delineation would affect classification success. The descriptive statistics for Random Forest from IWS crowns are presented in Table 1.

The descriptive statistics for Random Forest from crowns delineated using Detectree2 are presented in Table 2. The results indicate that the impact of delineation on classification accuracy is minimal, and both outputs provide similar accuracy. The YOLO model correctly

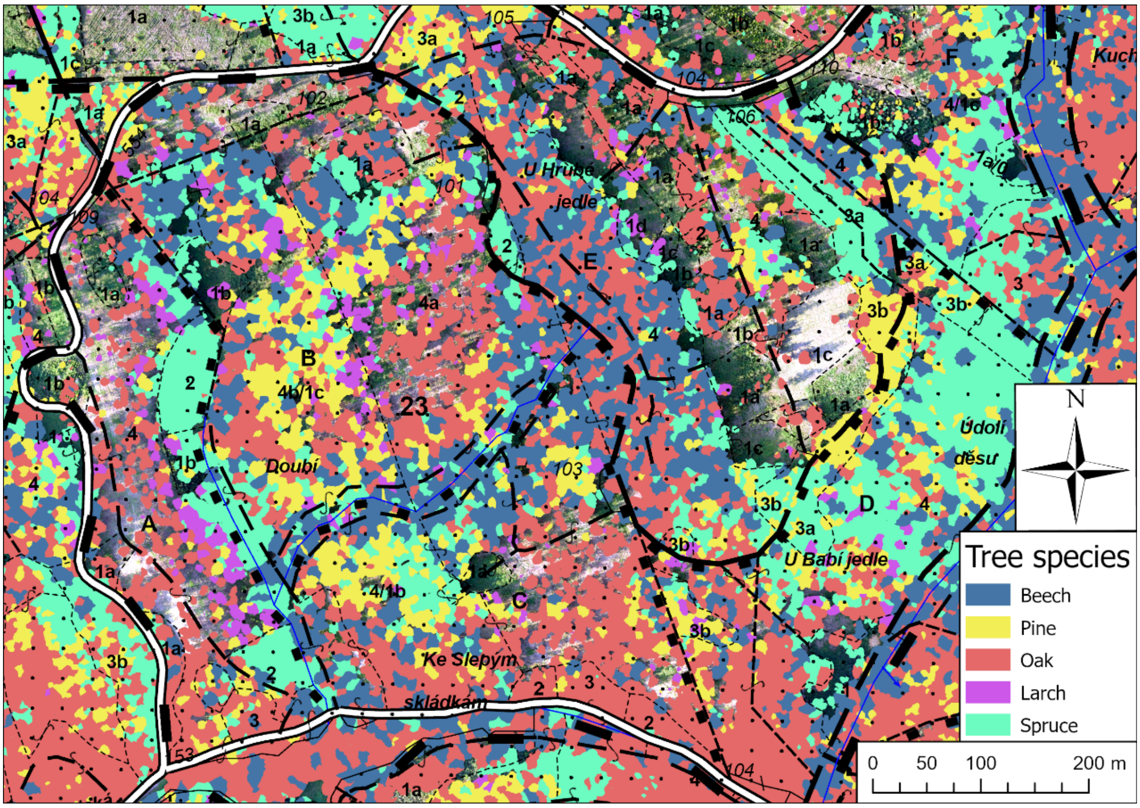

Fig. 2: *Map of tree species created from UAV data using inverse watershed segmentation and Random Forest algorithm*

Category	F1- score	Recall	Precision	Producer‘s Accuracy	User‘s Accuracy
Oak	0,6	0,5	0,75	0,866	0,962
Spruce	0,736	0,636	0,875	0,923	0,888
Pine	0,533	0,666	0,444	0,95	0,826
Beech	0,666	0,833	0,555	0,806	0,925
Larch	0,545	0,6	0,5	0,916	0,814

Tab. 2 *Evaluation of the classification performed by Random Forest and Detectree2*
Kappa Index: 0,856
Overall Accuracy: 0,885

classified 70% of spruces, with minor error rates for other tree categories and the background. 69% of larches were correctly classified, but 36% were mistakenly labeled as background. Only 39% of pines were correctly classified, with 27% misclassified as deadwood and 27% as background, indicating significant classification errors. 47% of deadwood was correctly classified, while 53% was incorrectly identified as background. Background was classified very well with 61% accuracy, but errors occurred in classifying spruce (26%), larch (28%), and deadwood (9%) as background.

In case of satellite data, the overall accuracies obtained were 76.1% for Sentinel-2 and 71% for PlanetScope, with Kappa coefficients of 0.704 and 0.642 and The Average F1 score of 74.29% and 67.7%, respectively. Overall scores of predictions are 3% percent lower for Sentinel-2 and 8% lower than accuracy of their models. These results are informative when placed in the context of similar research efforts, which have utilized both similar anddiffering methodologies.

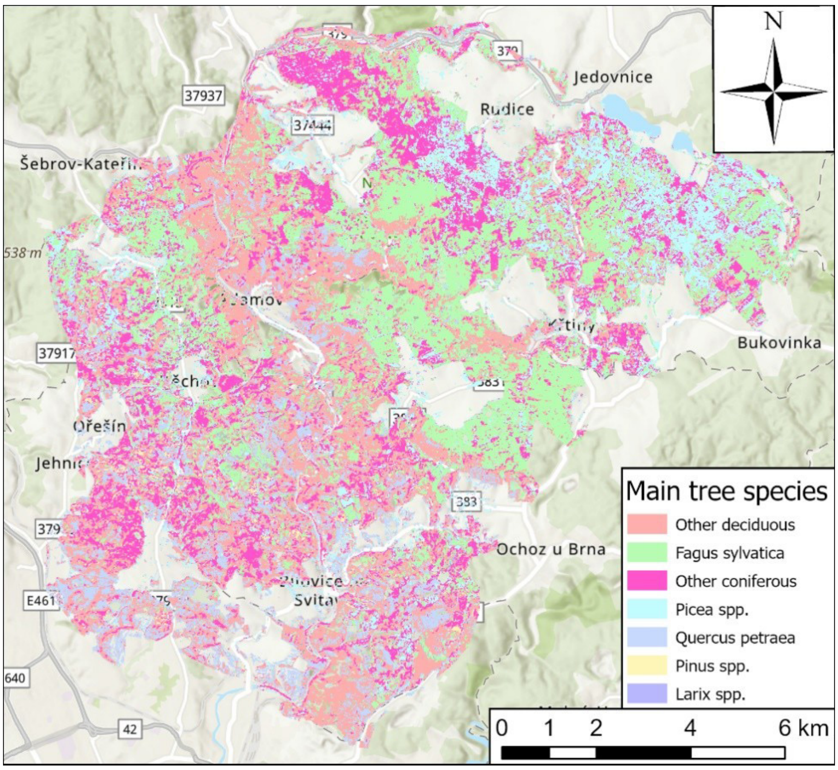

Fig. 3: *Main forest types (tree species) at the University Forest Enterprise Křtiny*

4 Discussion

The YOLO algorithm demonstrated its ability to accurately identify spruces, but the overall detection completeness was low, which could lead to certain tree species being overlooked. Despite the algorithm being trained on almost half of the area of interest, its overall completeness remained low. This result points to several potential issues. The main issue is the insufficient size of the training dataset. Studies that specifically address the size of datasets needed for training robust algorithms like YOLO highlight that the dataset should consist of thousands of images (Terven et Cordova-Esparza, 2023). Another potential problem is the use of natural data. While this presents certain challenges in deep learning, it doesn‘t necessarily lead to incorrect algorithm training. Natural data are often complex and variable, which can be challenging for models because they may contain noise, variability, and inconsistencies that are uncommon in controlled or synthetic datasets. The complexity and variability of natural data can drive algorithms to become more robust, as they must learn to extract relevant patterns and features from a challenging data environment, but this requires a sufficiently large dataset (Nat Biotechnol., 2023). The YOLO algorithm was trained only on RGB data, which put it at a disadvantage compared to Random Forest. Random Forest was trained on RGB data, multispectral data, vegetation indices, and 3D data in the form of a crown height model. Some improvement could be achieved by using a dataset that combines RGB and multispectral data. This would provide the algorithm with sufficient information, potentially enabling it to achieve results comparable to Random Forest. The results could also be improved by incorporating so-called point cloud metrics, which describe the distribution of points in the point cloud and possibly their color among the predictors. Point cloud metrics, also known as lidar metrics, are widely used in the forest inventory method called ABA (Area-based approach). The algorithm would then also work with the shape of the crowns (White *et al.*, 2013).

In case of satellite data analysys, the overall accuracy can be improved by a better set of tree sample plots. Sample plots that were used in the study were not fully precise, and it could cause higher misclassification. The Forest inventory was performed without fixation of coordinates on differential global navigation satellite system (GNSS) device but on a low accuracy level device. Incorporation of texture features (mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment, and correlation of Principal components of imagery) in addition to sentinel-2 imagery and topographic features in (Ma *et al.*, 2021) study had an overall accuracy of 86.49% and a Kappa of 0.83. This suggests that integrating textural information could potentially enhance the accuracy of classifications. CHM was the most important variable in both models confirming the conclusions of other studies, where topographic features were causing significant improvements in classification. The temporal features also have a great value for classification. Spectral indices of winter images were frequently identified among the most important variables. This possibly could be due to some changes in phenological periods causing shifts in the spectral response of trees, enhancing the visibility of differences between species for machine learning classifiers. Study of (Kluczek *et al.*, 2023) suggests that spring and autumn are the best sampling periods in context of species classification.

5 Summary

The Random Forest model was trained using both types of crowns, first with those delineated by the IWS method and then with crowns delineated by Detectree2. Although the methods differed in crown segmentation accuracy, both approaches achieved very good overall accuracy in tree species classification, demonstrating the robustness of the Random Forest algorithm against variability in input data. The overall accuracy of over 88% is a significant achievement, highlighting the effectiveness of the combined approach in tree species classification. This result not only confirms the importance of careful selection of crown segmentation methods but also emphasizes the potential of integrating various techniques to improve classification models.

YOLO, despite its success in identifying spruce, faced gaps in overall detection completeness, indicating the need for expanded training and validation datasets to improve the model. These findings underscore the importance of sufficiently large and representative datasets for training robust models in natural data environments. Given the combination of using multiple methods for tree species detection, it can be concluded that for the area of interest, the Random Forest algorithm is more suitable, as it could work with more information about the specific tree. Issues may have arisen in connecting crowns to labeled trees, highlighting the importance of precise crown delineation for classifier training. It was found that shortcomings in the training dataset, including ambiguous crown identification from aerial imagery, may have negatively impacted model performance.

Despite not achieving the accuracy levels noted in other studies regarding the satellite data analysis, the results of this research are promising and provide valuable insights into effective strategies for enhancing classification accuracy in future. This study underscores the importance of dataset composition and demonstrates the significant potential of integrating Random Forest (RF) algorithms with Sentinel-2 imagery and Canopy Height Model for efficient tree species mapping. This method of producing high-quality tree species maps at an effective cost can significantly enhance environmental monitoring efforts and serve as a valuable tool for sustainable forest management and conservation.

References

KLUCZEK, Marcin, ZAGAJEWSKI, Bogdan, ZWIJACZ-KOZICA, Tomasz, 2023. Mountain Tree Species Mapping Using Sentinel-2, PlanetScope, and Airborne HySpex Hyperspectral Imagery. *Remote Sensing*. 15, 3, 884. https://doi.org/10.3390/rs15030844

MA, Minfei *et al*. 2021. Tree species classification based on sentinel-2 imagery and random forest classifier in the eastern regions of the qilian mountains. *Forests*. 12, 12, 1736. https://doi.org/10.3390/f12121736

NAT BIOTECHNOL. 2023. *Data sharing in the age of deep learning*. https://doi.org/10.1038/s41587-023-01770-3

WHITE, J. C., WULDER, M. A., VARHOLA, A., VASTARANTA, M., COOPS, N. C.,COOK, B. D., PITT, D., WOODS, M. 2013. *A best practices guide for generating forest inventory attributes from airborne laser scanning data using the area-based approach*. Information Report FI-X-10. Natural Resources Canada, Canadian Forest Service, Canadian Wood Fibre Centre, Pacific Forestry Centre, Victoria, BC. 50 p. http://cfs.nrcan.gc.ca/publications?id%26hairsp%3B=%26hairsp%3B34887

TERVEN, J., CORDOVA-ESPARZA, D. 2023. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*. 5, 1680–1716. https://doi.org/10.48550/arXiv.2304.00501

Acknowledgement

This research was supported by the project „Sowing using unmanned aerial vehicles and biodegradable capsules“, Nr. IGA24-FFWT-TP-002, within Internal Grant Agency of FFWT Mendelu in Brno.

Contact information

Zdeněk Patočka: e-mail: zdenek.patocka@mendelu.cz