Multi-Temporal Tree Species Classification using Sentinel-2 Imagery and Machine Learning: A Case Study from Bolu, Aladağ Forests

Arif Aksüt [1], Semih Kuter [2]

[1] Çankırı Karatekin University, Graduate School of Natural and Applied Sciences, Dept. of Forest Engineering, Türkiye - arifaksutt@gmail.com
[2] Çankırı Karatekin University, Faculty of Forestry, Dept. of Forest Engineering, Türkiye - semihkuter@karatekin.edu.tr

Keywords: Sentinel-2, Machine Learning, Tree Species Classification, Feature Importance

Abstract

This study investigates the potential of multi-temporal Sentinel-2 imagery combined with machine learning techniques for tree species classification in the Aladağ Forests of Bolu province, Türkiye. Five dominant species—Black Pine, Scots Pine, Nordmann Fir, Beech, and Sessile Oak—were classified using a comprehensive feature set comprising spectral bands, spectral band indices, topographic attributes, and seasonal indicators. Sentinel-2 images acquired in April, August, and November were employed to capture phenological variations influencing classification accuracy. Field data from 112 sample plots supported the training and validation of four machine learning models: Random Forest (RF), Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Ensemble Learning (EL). On the independent test dataset, the overall classification accuracies were 99.22% for SVM with polynomial kernel, 99.03% for SVM with Radial Basis Function (RBF) kernel, 98.61% for RF, 97.91% for EL, and 96.66% for ANN, respectively. Seasonal analysis showed that August imagery provided the best classification performance, benefiting from peak canopy contrast, while accuracy decreased in November, particularly for deciduous species. The results underline the effectiveness of RF and emphasize the importance of integrating multi-seasonal satellite observations with machine learning for improved forest species mapping.

1. Introduction

Forests are among the most vital natural resources, possessing both exhaustible and renewable characteristics. With the increasing environmental challenges of today, the importance of forests has become more evident than ever before. In particular, global warming and climate change have brought forest ecosystems to the focus of research and conservation efforts. Forests play a critical role not only in preserving biodiversity but also in regulating the climate as carbon sinks. They also provide essential ecosystem services such as managing the water cycle, preventing soil erosion, and offering habitats for various species (Marta *et al.*, 2021).

In recent years, remote sensing technologies have been widely used to support the sustainable management and monitoring of forest resources. Traditional forest inventory methods are often costly, time-consuming, and limited to small areas, making these new technologies increasingly valuable. With the advancements in satellite systems and data processing algorithms, it has become possible to monitor large, forested areas both spatially and temporally. Sentinel-2 satellites, developed by the European Space Agency (ESA), are commonly used in remote sensing studies due to their medium- to-high spatial resolution, wide spectral range, and free accessibility (Drusch *et al.*, 2012). Sentinel-2 data play an important role in vegetation analysis as well as in the assessment of soil moisture, water bodies, and other environmental indicators.

These developments in remote sensing have also required the advancement of data processing and interpretation methods. In this context, machine learning algorithms have shown significant success, particularly in classification and modelling tasks. Recent studies have demonstrated that machine learning algorithms such as RF, SVM, EL methods, and ANN provide high classification accuracy in forest mapping (Freudenberg *et al.*, 2025; Hościło and Lewandowska, 2019; Liu *et al.*, 2024; Wessel *et al.*, 2018). These algorithms can handle high-dimensional and multivariate datasets and offer strong modelling capabilities beyond classical statistical methods. As a result, classical classification approaches such as Minimum Distance Classifier (MDC) and Maximum Likelihood Classification (MLC) are increasingly being replaced by these advanced machine learning techniques.

Improving classification accuracy in forest studies depends not only on the choice of algorithm but also on the quality and timing of the variables used. Seasonal variability, especially in deciduous tree species, can directly affect spectral reflectance characteristics and thus influence classification performance (Liu *et al.*, 2024). Therefore, comparing data collected in different periods of the year is an important step to improve model performance.

The main objective of this study is to classify five different tree species located in the Bolu, Aladağ Forest Management Directorate using Sentinel-2 satellite data. Initially, a large set of vegetation, water, soil indices, and topographic variables was generated from the remote sensing data.

During data analysis and classification, machine learning algorithms including Random Forest (RF), Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Ensemble Learning (EL) were applied. The performance of each algorithm was evaluated using different parameter combinations, and the results were analysed based on accuracy metrics derived from the associated error matrices. In addition, the effect of seasonal variation on classification performance was tested, and the highest accuracy was obtained during the summer months.

2. Materials and Methods

2.1 The Study Area

The study area covers 88,717.05 hectares (approximately 887 km²) and is located within the Bolu, Aladağ Forest Management Directorate (cf. Figure 1). The dominant tree species include the conifers Black Pine (*Pinus nigra*), Scots Pine (*Pinus sylvestris*), and Nordmann Fir (*Abies nordmanniana* ssp.), as well as the broadleaf species Beech (*Fagus* spp.) and Sessile Oak (*Quercus petraea*). The total area consists of 54,165.7 hectares of coniferous stands and 2,840.2 hectares of broadleaf stands, resulting in a combined total of 57,005.9 hectares of forested land (OGM, 2018) (cf. Table 1). The region has a humid continental climate, characterized by cold, snowy winters and warm summers. Annual precipitation is relatively high, supporting dense forest cover, while seasonal variability influences vegetation dynamics and phenological stages, which are important for remote sensing-based forest monitoring (Kantarcı, 1980).

Tree Species	**Area (ha)**
Black Pine	10,052.4
Scots Pine	24,577.8
Nordmann Fir	19,535.5
Beech	997.2
Sessile Oak	1,843.0
Total	**57,005.9**

Table 1. Summary of the main tree species composition in the study area.

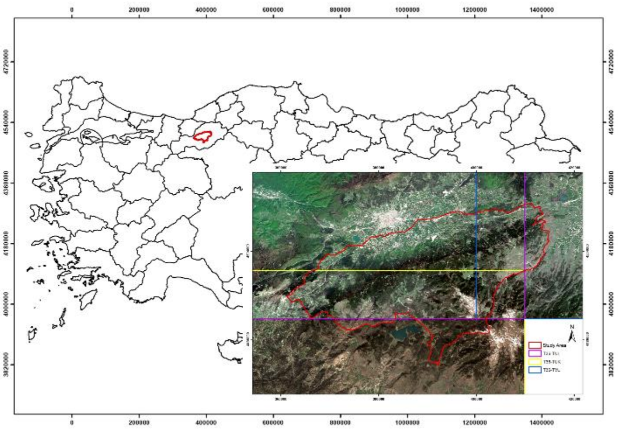

Figure 1. Study area and the locations of Sentinel-2 tiles.

2.2 Sentinel-2 Image Dataset

Sentinel-2 is a multispectral optical satellite mission developed by the European Space Agency (ESA) within the Copernicus Programme. The mission consists of two identical satellites (Sentinel-2A and Sentinel-2B) operating in a sun-synchronous orbit, providing global coverage with a revisit time of 5 days at the equator (Drusch *et al.*, 2012). Sentinel-2 delivers imagery across 13 spectral bands ranging from visible to shortwave infrared (443–2190 nm), with spatial resolutions of 10 m, 20 m, and 60 m depending on the band. Its wide swath width of 290 km enables systematic monitoring of terrestrial ecosystems, vegetation, land cover, and forest dynamics. The high temporal, spectral, and spatial resolution of Sentinel-2 makes it particularly well-suited for forest classification, phenological analysis, and environmental monitoring applications. In this study, Sentinel-2 imagery was utilized as the primary remote sensing data source to support multi-temporal forest species classification.

The study area overlaps with three Sentinel-2 tiles (T36TUL, T36TUK, and T36TVL) (cf. Figure 1). A total of 17 Sentinel-2 scenes acquired in 2018 were downloaded from the Copernicus Open Access Hub (https://www.copernicus.eu/). To capture key phenological stages, images from April (leaf onset), August (peak canopy development), and November (leaf senescence) were selected, following the approach of Liu *et al.* (2023). The Sentinel-2 Level-1C images, with a spatial resolution of 20 meters, were atmospherically corrected using the Sen2Cor processor to generate Level-2A surface reflectance products. This preprocessing step minimized atmospheric distortions and ensured radiometric consistency across the selected dates, providing reliable input data for the classification analysis.

In this study, a field survey was conducted between 2-10 September 2023 to collect reference data for model training and accuracy assessment. The sampling design considered factors influencing classification performance, including tree species composition, canopy structure, and species mixture ratios. A 500-meter buffer zone (250 meters on each side) was generated along the existing road network to optimize field accessibility. A systematic grid was established at 500-meter intervals, generating 1120 candidate sampling points. From these, 162 points were randomly selected based on the areal distribution of tree species and canopy closure conditions.

Due to the close proximity of some points within homogeneous stand types and limited road access in certain locations, final field measurements were successfully collected at 112 sampling points. From these 112 sampling points, 75 were identified as pure stands and 37 as mixed stands. Based on the field plot records and random selection, sample sites were assigned for each tree species: 11 plots for Black Pine, 20 for Scots Pine, 10 for Nordmann Fir, 1 for Beech, and 2 for Sessile Oak. Considering factors such as data homogeneity, class purity, and accurate spatial matching, a subset of 44 plots representing the five tree species was selected from the initial 112 plots for model training (cf. Figure 2).

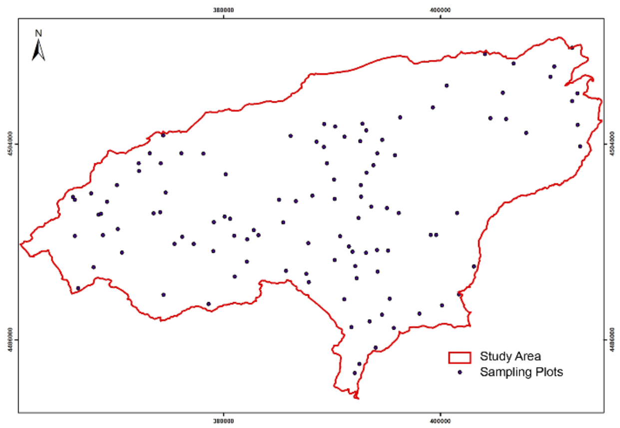

Figure 2. The locations of 112 random sampling plots.

The final dataset was assembled by integrating multiple feature groups derived from Sentinel-2 imagery and auxiliary data sources. Specifically, 9 spectral bands, 12 vegetation, water, and soil indices computed from these bands, and 4 principal components obtained through Principal Component Analysis (PCA) were included. In addition, elevation and aspect data extracted from the digital elevation model were incorporated, resulting in 27 continuous predictor variables. To account for seasonal variability, month information (April, August, November) was added as a categorical variable, increasing the total number of input features to 28.

The target variable was defined according to tree species classes: Black Pine ("1"), Scots Pine ("2"), Nordmann Fir ("3"), Beech ("4"), and Sessile Oak ("5"). Thus, the complete dataset comprised 29 columns combining both predictors and target classes. The dataset was randomly partitioned into training (70%) and testing (30%) subsets using MATLAB to facilitate model calibration and performance assessment.

2.3 Machine Learning Algorithms

Four machine learning algorithms were employed to perform tree species classification: RF, SVM, ANN and tree- based EL methods. The RF algorithm constructs multiple decision trees using bootstrap sampling and random feature selection at each node, and aggregates the results through majority voting to improve classification stability and reduce overfitting (Breiman, 2001). SVM transforms the input data into a higher-dimensional feature space using kernel functions and identifies the optimal separating hyperplane that maximizes the margin between classes (Vapnik *et al.*, 1997). ANN consists of interconnected layers of neurons with adaptive weights and non-linear activation functions, enabling the model to capture complex, non-linear relationships in the data (Basheer and Hajmeer, 2000). The tree-based EL approach combines multiple decision trees through boosting or bagging techniques, aggregating the predictions of multiple weak learners to achieve improved generalization and predictive accuracy (Zhang and Ma, 2012). All models were trained and validated on identical training and test datasets to enable a consistent and fair comparison of their classification performances.

Following the initial classification, misclassifications and noise artifacts can negatively affect both the accuracy and the spatial coherence of the results. To reduce such errors, post- classification spatial filtering techniques were applied. In this study, all classification outputs were produced in five classes and subsequently processed using a 5×5 Majority Filter. This filter assigns each pixel the most frequently occurring class value within its defined neighbourhood window, thereby smoothing isolated misclassified pixels and enhancing spatial consistency.

3. Results and Discussions

In this study, classification performance was assessed in detail for the RF, SVM, ANN, and EL algorithms using multiple accuracy metrics and confusion matrices. For each algorithm, standard evaluation measures such as Overall Accuracy (OA), Producer's Accuracy (PA), and User's Accuracy (UA) were calculated based on both training and test datasets. The results indicated that all machine learning algorithms achieved higher overall accuracies on the training data, while a slight decrease in accuracy was observed on the test data, reflecting the models' generalization capabilities. Confusion matrix analyses revealed that misclassifications were more frequent among spectrally similar species, and the degree of confusion varied depending on the algorithm applied. Overall, RF demonstrated superior performance, particularly in cases where species showed distinct spectral and structural characteristics. These findings emphasize the importance of selecting an appropriate classification algorithm that aligns with the specific data structure and class separability characteristics of the study area. The RF classification results on the test dataset is given in Table 2.

		Reference					
		B.P.	**S.P.**	**N.F.**	**B.**	**S.O.**	**Total**
Classified	**B.P.**	790	0	14	1	0	805
	S.P.	7	962	4	0	0	973
	N.P.	4	6	705	0	0	715
	B.	0	0	0	62	0	62
	S.O.	0	0	0	1	121	122
	Total	801	968	723	64	121	**2677**

Table 2. RF classification error matrix on test data (B.P.: Black Pine; S.P.: Scots Pine, N.P.: Nordmann Fir; B.: Beech; S.O.: Sessile Oak)

The classification performance of the RF model was further analyzed using detailed accuracy metrics derived from the independent test dataset. The PA values were 98.63% for Black Pine *(Pinus nigra*), 99.38% for Scots Pine (*Pinus sylvestris*), 97.67% for Nordmann Fir (*Abies nordmanniana* ssp.), 96.87% for Beech (*Fagus* spp.), and 100% for Sessile Oak (*Quercus petraea*). Correspondingly, the UA values were 98.25% for Black Pine, 98.86% for Scots Pine, 98.60% for Nordmann Fir, 100.00% for Beech and 99.18% for Sessile Oak.

The OA achieved by the RF model on the test dataset reached 98.65%, confirming the high classification reliability and generalization capability of the model across both coniferous and broadleaf species present in the study area (cf. Figure 3).

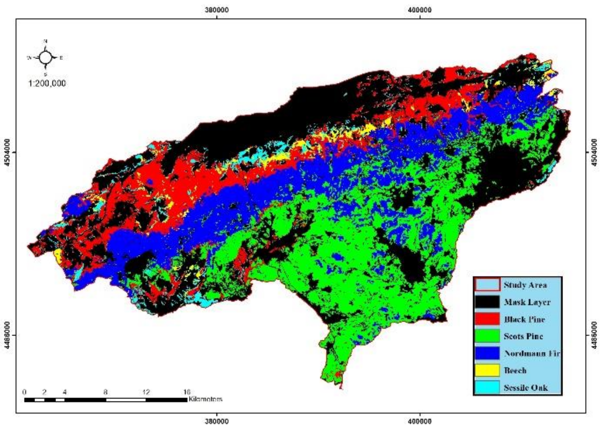

Figure 3. The RF classification on August for the study area (Note: The mask layer is derived from the forest management map of the study area and it includes non-forested areas such as agriculture, built-up, etc.).

		Reference					
		B.P.	**S.P.**	**N.F.**	**B.**	**S.O.**	**Total**
Classified	**B.P.**	797	1	7	0	0	805
	S.P.	1	964	4	0	0	969
	N.P.	2	2	712	0	0	716
	B.	0	0	0	63	1	64
	S.O.	1	1	0	1	120	123
	Total	801	968	723	64	121	**2677**

Table 3. SVM- Polynomial Kernel classification error matrix on test data (B.P.: Black Pine; S.P.: Scots Pine, N.P.: Nordmann Fir; B.: Beech; S.O.: Sessile Oak)

The classification performance of the SVM model with the polynomial kernel was examined using the test dataset. The PA values obtained were 99.50% for Black Pine (*Pinus nigra*), 99.59% for Scots Pine (*Pinus sylvestris*), 98.48% for Nordmann Fir (*Abies nordmanniana* ssp.), 98.44% for Beech (*Fagus* spp.), and 99.17% for Sessile Oak (*Quercus petraea*). The corresponding UA values were calculated as 99.01% for Black Pine, 99.48% for Scots Pine, 99.44% for Nordmann Fir, 98.44% for Beech and 97.56% for Sessile Oak. The SVM model employing a polynomial kernel achieved an OA of 99.22%.

		Reference					
		B.P.	**S.P.**	**N.F.**	**B.**	**S.O.**	**Total**
Classified	**B.P.**	796	1	8	0	0	805
	S.P.	2	962	5	0	0	969
	N.P.	3	2	710	0	0	715
	B.	0	0	0	63	1	64
	S.O.	0	3	0	1	120	124
	Total	801	968	723	64	121	**2677**

Table 4. SVM- RBF Kernel classification error matrix on test data (B.P.: Black Pine; S.P.: Scots Pine, N.P.: Nordmann Fir; B.: Beech; S.O.: Sessile Oak).

The total number of validation instances used in this evaluation was slightly similiar from those used in other models. The performance of the SVM model utilizing an RBF kernel was assessed on the independent test dataset. The PA values were calculated as 99.38% for Black Pine (*Pinus nigra*), 99.38% for Scots Pine (*Pinus sylvestris*), 98.20% for Nordmann Fir (*Abies nordmanniana* ssp.), 98.44% for Beech (*Fagus* spp.), and 99.17% for Sessile Oak (*Quercus petraea*). Correspondingly, the UA values were 98.88% for Black Pine, 99.28% for Scots Pine, 99.30% for Nordmann Fir, 98.44% for Beech and 96.77% for Sessile Oak. The OA test was calculated for SVM- RBF kernel 99.03%.

		Reference					
		B.P.	**S.P.**	**N.F.**	**B.**	**S.O.**	**Total**
Classified	**B.P.**	791	0	14	0	0	805
	S.P.	3	954	9	0	0	966
	N.P.	5	13	700	0	0	718
	B.	0	0	0	60	0	60
	S.O.	2	1	0	4	121	128
	Total	801	968	723	64	121	**2677**

Table 5. ANN classification error matrix on test data (B.P.: Black Pine; S.P.: Scots Pine, N.P.: Nordmann Fir; B.: Beech; S.O.: Sessile Oak).

The performance of the ANN model was assessed using a test dataset., and its classification performance was quantified using class-specific accuracy metrics. The PA was calculated as 98.75% for Black Pine (*Pinus nigra*), 98.55% for Scots Pine (*Pinus sylvestris*), 96.82% for Nordmann Fir (*Abies nordmanniana* ssp.), 93.75% for Beech (*Fagus* spp.) and 100% for Sessile Oak (*Quercus petraea*). The UA values were found to be 98.26% for Black Pine, 98.76% for Scots Pine, 97.49% for Nordmann Fir, 100% for Beech and 94.53% for Sessile Oak. The OA achieved by the ANN model on the test dataset reached 98.09%.

		Reference					
		B.P.	**S.P.**	**N.F.**	**B.**	**S.O.**	**Total**
Classified	**B.P.**	787	2	15	3	0	807
	S.P.	6	939	13	0	0	958
	N.P.	7	27	695	0	0	729
	B.	0	0	0	60	0	60
	S.O.	1	0	0	1	121	123
	Total	801	968	723	64	121	**2677**

Table 6. EL classification error matrix on test data (B.P.: Black Pine; S.P.: Scots Pine, N.P.: Nordmann Fir; B.: Beech; S.O.: Sessile Oak).

The EL model, which integrates multiple base classifiers to improve prediction robustness, was evaluated on the independent test dataset. The classification performance demonstrated high accuracy across all species. The PA values were 98.25% for Black Pine (*Pinus nigra*), 97.00% for Scots Pine (*Pinus sylvestris*), 96.13% for Nordmann Fir (*Abies nordmanniana* ssp.), 93.75% for Beech (*Fagus* spp.) and 100.00% for Sessile Oak (*Quercus petraea*). Similarly, the UA values were 97.52% for Black Pine, 98.02% for Scots Pine, 95.34% for Nordmann Fir 100.00% for Beech and 98.37% for Sessile Oak. The OA of the EL model was computed as 97.20%.

The seasonal performance analysis revealed notable variations in classification accuracy across tree species and phenological periods. For Black Pine (*Pinus nigra*), accuracy decreased slightly from 82.46% in April to 81.00% in August, followed by a modest increase to 82.37% in November, indicating relatively stable spectral separability throughout the year. Scots Pine (*Pinus sylvestris*) and Nordmann Fir (*Abies nordmanniana* ssp.) exhibited more stable classification results, maintaining high accuracy even in November (92.52%) for Scots Pine and in August (76.98%) for Nordmann Fir, suggesting that these species are less affected by phenological changes. Black Pine, Scots Pine, and Nordmann Fir demonstrated consistent classification performance across all months, indicating its spectral stability and suitability for year-round mapping under the conditions of this study. Overall, coniferous species were minimally affected by phenological changes.

In the case of Beech (*Fagus* spp.), accuracy was extremely low in April (18.29%) but increased substantially to 84.56% in August. This sharp seasonal contrast suggests that Beech develops more distinctive spectral characteristics during full leaf development in summer, while becoming nearly indistinguishable from other species during leaf-off periods, as reflected by 61.83% accuracy in November. For Sessile Oak (*Quercus petraea*), accuracy declined gradually from 73.73% in April to 70.85% in August, and further to 66.82% in November, indicating increasing classification challenges towards the end of the growing season. These patterns may be associated with changes in leaf coloration, moisture loss, and declining photosynthetic activity.

Overall, August emerged as the most favourable period for tree species classification, providing the highest accuracies across all species due to enhanced spectral separability during peak vegetation activity. In contrast, November showed a marked decrease in classification performance for deciduous species, primarily due to increased spectral similarity and canopy senescence.

4. Conclusions

This study demonstrated the effective use of Sentinel-2 imagery combined with machine learning algorithms for tree species classification in the Bolu, Aladağ Forest Management Directorate. Four classification models RF, SVM, ANN, and EL were applied and systematically compared. Among the models tested, RF exhibited superior performance with an OA of 98.65% on the test data. This result slightly surpasses the 96% accuracy reported by Hemmerling *et al*. (2021) and is also higher than the 92.38% achieved by Grabska *et al*. (2019) in classifying nine tree species using Sentinel-2 imagery. This difference may be due to the use of multi-seasonal imagery and a more diverse set of vegetation indices in the current study. While both studies used Sentinel-2 time-series data and RF algorithms, the higher accuracy in this study may be attributed to the inclusion of a broader set of vegetation indices and careful sample balancing strategies. Seasonal analysis further highlighted the influence of phenological variability on classification accuracy, with August emerging as the optimal period for species discrimination due to enhanced spectral separability during peak vegetation activity.

The results emphasize the importance of integrating multi-temporal remote sensing data with robust classification algorithms to improve species-level forest mapping. The proposed approach provides a reliable framework for operational forest monitoring and management in support of sustainable forestry practices.

Acknowledgements

This study was carried out as part of the Ph.D. research of Arif Aksüt, conducted at the Çankırı Karatekin University, Graduate School of Natural and Applied Sciences, Department of Forest Engineering. The authors would like to express their gratitude to the Bolu, Aladağ Forest Management Directorate for their valuable support and assistance during the field data collection phase.

References

Basheer, I. A. and Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods, 43*(1), pp. 3-31.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), pp. 5-32.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F. and Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment, 120*, pp. 25-36.

Freudenberg, M., Schnell, S. and Magdon, P. (2025). A Sentinel-2 machine learning dataset for tree species classification in Germany. *Earth Syst. Sci. Data, 17*(2), pp. 351-367.

Hemmerling, J., Pflugmacher, D. and Hostert, P. (2021). Mapping temperate forest tree species using dense sentinel-2 time series. *Remote Sensing of Environment, 267*, p. 112743. doi:10.1016/j.rse.2021.112743.

Hościło, A. and Lewandowska, A. (2019). Mapping Forest Type and Tree Species on a Regional Scale Using Multi-Temporal Sentinel-2 Data. *Remote Sensing, 11*(8). doi:10.3390/rs11080929.

Grabska, E., Hostert, P., Pflugmacher, D. and Ostapowicz, K. (2019). Forest stand species mapping using the sentinel-2 time series. Remote Sensing, 11(10), p. 1197. doi:10.3390/rs11101197.

Kantarcı, D. (1980). Aladağ kütlesinin (Bolu) kuzey yamacında Uludağ Göknarı ibrelerindeki mineral madde miktarlarının yükselti-iklim kuşaklarına göre değişimi. *İstanbul Üniversitesi Orman Fakültesi Dergisi,* 30(2), pp. 135-145.

Liu, P., Ren, C., Wang, Z., Jia, M., Yu, W., Ren, H. and Xia, C. (2024). Evaluating the Potential of Sentinel-2 Time Series Imagery and Machine Learning for Tree Species Classification in a Mountainous Forest. *Remote Sensing, 16*(2). doi:10.3390/rs16020293

Liu, X., Frey, J., Munteanu, C., Still, N. and Koch, B. (2023). Mapping tree species diversity in temperate montane forests using Sentinel-1 and Sentinel-2 imagery and topography data. *Remote Sensing of Environment,* 292, pp. 113576.

Marta, S., Brunetti, M., Manenti, R., Provenzale, A. and Ficetola, G. F. (2021). Climate and land-use changes drive biodiversity turnover in arthropod assemblages over 150 years. *Nature Ecology & Evolution,* 5(9), pp. 1291-1300.

OGM. (2018). Orman Genel Müdürlüğü Ekosistem Tabanlı Fonksiyonel Orman Amenajman Planı.

Vapnik, V., Golowich, S. E. and Smola, A. J. (1997). Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. In M. C. Mozer, M. I. Jordan and T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9* (pp. 281-287): MIT Press.

Wessel, M., Brandmeier, M. and Tiede, D. (2018). Evaluation of Different Machine Learning Algorithms for Scalable Classification of Tree Types and Tree Species Based on Sentinel-2 Data. *Remote Sensing, 10*(9). doi:10.3390/rs10091419

Zhang, C. and Ma, Y. (Eds.). (2012). *Ensemble Machine Learning: Methods and Applications*: Springer.