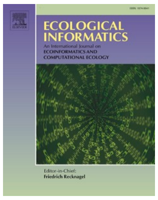
Ecological Informatics 87 (2025) 103133

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Validation of sentinel 2 based machine learning models for Czech National Forest Inventory

Richard Kovárník [*], Jitka Janová

Department of Statistics and Operation Analysis, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 61300 Brno, Czech Republic

ARTICLE INFO

Keywords:
Machine learning
Remote sensing
Data science
Forest management
National forest inventory
Automation

ABSTRACT

The National Forest Inventory (NFI) of the Czech Republic provides essential data for forest management but requires significant time and resources. This study highlights the critical role of validating Sentinel-2-based machine learning models against real NFI data to ensure their reliability for forest monitoring. While satellite-based models offer a cost-effective alternative, their practical applicability depends on rigorous validation. We applied four commonly used machine learning models—Classification and Regression Trees, Random Forest, Support Vector Machine, and Naive Bayes—to Sentinel-2 imagery to estimate forest cover conditions. The Random Forest model achieved the highest overall accuracy (98.3 %). By systematically comparing model predictions with official NFI data, we address a key gap in remote sensing applications: the need for real-world validation beyond training datasets. Our findings demonstrate that properly validated Sentinel-2-based models can enhance large-scale forest monitoring, reducing the financial and labor burdens of traditional field surveys while ensuring data accuracy for sustainable forest management.

1. Introduction

The accelerating pace of urbanization and environmental change necessitates the development and application of advanced technologies to monitor landscape conditions and dynamics, thereby facilitating prompt and effective decision-making (Panahi et al., 2024). Remote sensing, machine learning, and geospatial analysis have emerged as crucial tools for tracking ecological changes, particularly in the field of forestry. These technologies enhance our ability to conduct large-scale environmental assessments with high precision and efficiency, reducing the reliance on traditional field-based data collection methods (Fassnacht et al., 2024).

In the Czech Republic, the National Forest Inventory (NFI), established under Forest Act 289/1995, serves as the primary method for systematically collecting data on forest stand conditions at the national level (Institute for Forest Management, 2024). Initiated as a national project in 2001, the NFI employs field surveys supplemented by photogrammetry and digital aerial image analysis to monitor and assess forests across extensive areas. While this approach is valued for its accuracy and reliability, it faces challenges such as high costs, time-consuming data collection, limited update frequency, and the need for substantial personnel and technological resources (Kangas and Maltamo, 2006). Consequently, there is a growing interest in innovative technologies that can complement or partially replace traditional methods, offering faster and more cost-effective solutions (McRoberts and Tomppo, 2007).

The advancement of computer technology has led to the emergence of systems capable of significantly reducing costs, notably through the acquisition and analysis of satellite imagery. These images are now obtained in high quality and at regular intervals, enhancing their utility in monitoring landscape changes (Coops et al., 2023). Remote sensing data is increasingly employed in scientific studies addressing forest structure, biomass estimation, and land cover classification (Fassnacht et al., 2024). The integration of Earth observation data with machine learning techniques has been shown to improve the accuracy of forest monitoring, allowing for more frequent and precise assessments of forest dynamics (Kalinaki et al., 2023). Proof of the topicality of this topic are also projects such as the LANDSUPPORT project (Terribile et al., 2024), which was part of the EU Horizon 2020 project (European Commission, 2020).

One of the most widely used platforms for processing and analysing satellite imagery is Google Earth Engine (GEE) (Gorelick et al., 2017). GEE provides access to extensive archives of remote sensing data and facilitates large-scale spatial analyses. In forestry applications, GEE has

---

* Corresponding author.
E-mail addresses: xkovarn1@mendelu.cz (R. Kovárník), jitka.janova@mendelu.cz (J. Janová).

https://doi.org/10.1016/j.ecoinf.2025.103133
Received 6 January 2025; Received in revised form 2 April 2025; Accepted 2 April 2025

been utilized for monitoring forest loss due to urban expansion (Zurqani et al., 2019), analysing forest species composition through participatory mapping (Koskinen et al., 2019), and assessing forest degradation trends (Chen et al., 2021). Moreover, GEE has been applied beyond forestry, including investigations of agricultural land use (Savitha and Talari, 2023), grassland ecosystem monitoring (Pinna et al., 2024), and irrigation impact assessments (Gumma et al., 2023).

Forestry applications of remote sensing often rely on satellite systems such as Landsat 8 and Sentinel-2, both of which offer high spatial and temporal resolution (Amani et al., 2020). Sentinel-2, in particular, provides superior spectral and spatial resolution compared to Landsat, making it highly effective for detecting subtle changes in vegetation, such as seasonal variations and early signs of forest degradation (Pérez-Cutillas et al., 2023). However, despite the growing accessibility of these datasets, there remains a need for studies that validate the quality and reliability of predictions derived from remote sensing data, particularly in temperate and Central European forest ecosystems.

Recent research has emphasized the importance of integrating multiple data sources to enhance forest monitoring. For instance, Zurqani (2025) demonstrated the potential of combining multi-sensor satellite imagery and GEDI LiDAR data to estimate forest height and biomass, producing high-resolution maps for metropolitan France. Similarly, Bountos et al. (2023) introduced FoMo-Bench, a multi-modal, multi-scale, and multi-task benchmark designed to improve remote sensing foundation models for forest monitoring. These studies highlight the necessity of incorporating diverse datasets and advanced modelling techniques to refine forest monitoring methodologies.

Despite the progress in remote sensing applications, research on the integration of machine learning models with NFI data remains limited, particularly in the context of Central European forests. While studies have validated the application of machine learning for forest classification and biomass estimation in North America and Asia (Tamiminia et al., 2020), there is a gap in research focused on its application to European forestry datasets. Addressing this gap is critical for enhancing the effectiveness of forest inventories and developing more adaptive strategies for forest management in the face of climate change and increasing anthropogenic pressures.

The objective of this study is to validate the performance of machine learning models trained on remote sensing data by comparing their outputs with the official NFI results for the Czech Republic. Specifically, we aim to explore the potential of integrating satellite imagery and advanced computational techniques to optimize future forest inventories and monitoring efforts. By bridging the gap between remote sensing technologies and national forest inventories, this research contributes to the development of more efficient, scalable, and accurate forest assessment methodologies, which are essential for sustainable forest management and environmental conservation in an era of rapid ecological change.

2. Material and methods

2.1. Data collection

All data collection, computations, and data processing were conducted using the Google Earth Engine platform (Gorelick et al., 2017). The Weka library (Bouckaert et al., 2016) was utilized to implement machine learning algorithms, including Classification and Regression Tree (CART), Random Forest, Naive Bayes, and Support Vector Machines (SVM), within a JavaScript environment.

The initial step involved obtaining appropriate training data for model estimation. Since machine learning algorithms classify satellite imagery based on pixel-level spectral values, the data had to be manually collected and labelled. A total of 2800 geographic points were selected and categorized into four distinct classes, with 700 points assigned to each class: Crop, Forest, Construction, and Water. The training dataset was sourced from the vicinity of Brno, Czech Republic. For data collection, an image from the Copernicus program was employed, specifically from the Sentinel-2 mission. The training data distribution is illustrated in Fig. 1.

Sentinel-2 provides high-resolution multispectral imaging with a global revisit frequency of every five days. The spectral bands listed in Table 1 were utilized for classification into the specified classes. The final dataset comprises 2800 observations with seven spectral features, which were subsequently divided into training (70 %) and test (30 %) sets to facilitate model evaluation.

The 20 m bands were automatically resampled to 10 m using Google Earth Engine's default nearest-neighbor resampling method, ensuring consistency in spatial resolution across all input features. Another variable that was included in the classification is the Normalized Difference Vegetation Index (NDVI). NDVI is a widely used remote sensing index that provides a measure of vegetation health and density based on satellite image data. NDVI is calculated from visible and near-infrared light reflected by vegetation (Huang et al., 2021). This index is calculated for each pixel according to the following formula.

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

Box plots illustrating the distribution of spectral band values for each land cover class are presented in Fig. 2. The values have been converted to Top-of-Atmosphere (TOA) reflectance, which quantifies the fraction of sunlight reflected from the Earth's surface as detected by a satellite sensor. TOA reflectance is typically expressed as a dimensionless value ranging from 0 to 1. The trained models were subsequently applied to classify the median cloud-masked Sentinel-2 image of the Czech Republic (January 1 to December 30, 2023) depicted in Fig. 3.

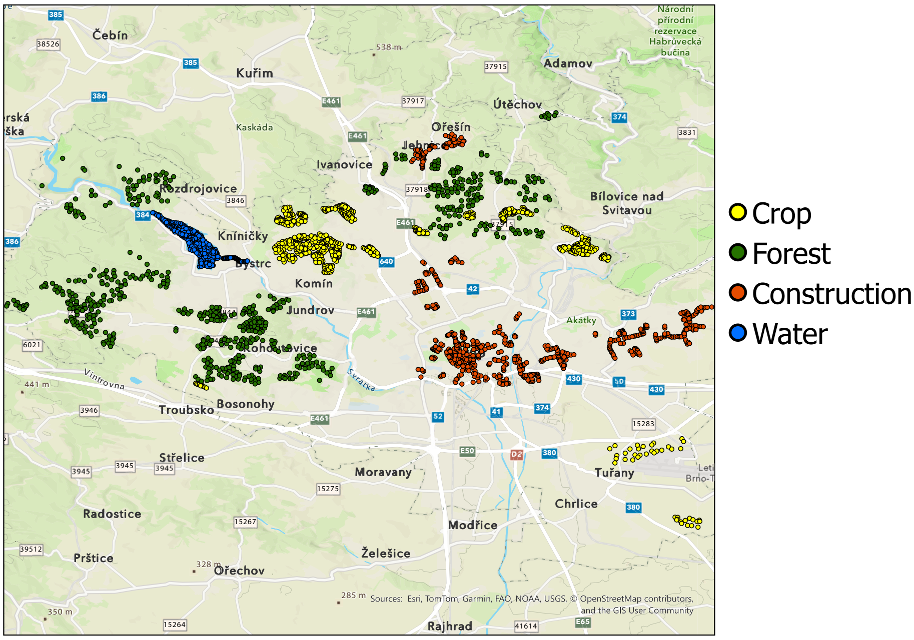
2.2. Methods

2.2.1. Supervised classification in GEE

A total of four supervised machine learning algorithms—Classification and Regression Trees (CART), Random Forest, Support Vector Machines (SVM), and Naive Bayes—were applied to the dataset using Google Earth Engine (GEE). These traditional machine learning methods are widely used in land cover classification due to their high accuracy, computational efficiency, and interpretability. The implementation of these models in GEE involved specific parameter adjustments to optimize their performance.

The CART algorithm, developed by Breiman (1996), is a widely used decision tree technique applicable to both classification and regression tasks. It constructs a binary decision tree where each node represents a decision rule based on a specific feature, and each branch leads to either an outcome or another decision node. In GEE, this model is implemented using ee.Classifier.smileCart() and includes two key parameters: maxNodes, which defines the maximum number of leaf nodes in each tree, and minLeafPopulation, which specifies the minimum number of samples required for a node to be split. By default, minLeafPopulation is set to 1. These parameters generally do not require extensive tuning but can be adjusted based on dataset characteristics and classification requirements (Gorelick et al., 2017).

Random Forest is an ensemble learning method that enhances decision tree performance by constructing multiple decision trees during training and aggregating their outputs. It determines the final class label based on the majority vote (classification) or average prediction (regression) of individual trees, thereby mitigating overfitting and improving model generalization (Genuer and Poggi, 2020). In GEE, the Random Forest classifier is implemented using ee.Classifier.smileRandomForest() and can be optimized by tuning several key parameters, including numberOfTrees (the total number of trees in the forest), variablesPerSplit (the number of features considered at each split), minLeafPopulation (the minimum number of samples per leaf node), bagFraction (the fraction of training data used per tree), and maxNodes (the maximum number of nodes per tree (Gorelick et al., 2017). The



Sources: Esri, TomTom, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community

Fig. 1. Training data collection.

Table 1
Sentinel 2 multispectral bands used for classification.

Band	Wavelength (nm)	Pixel Size (m)
B2 (Blue)	496.6	10
B3 (Green)	560	10
B4 (Red)	664.5	10
B8 (Near-Infrared)	835.1	10
B11 (Short-Wave Infrared 1)	1613.7	20
B12 (Short-Wave Infrared 2)	2202.4	20

algorithm was applied using the default parameter values, with the exception of the parameter determining the number of trees. Default values for the other parameters generally provide sufficient performance for most remote sensing tasks, eliminating the need for extensive hyperparameter tuning.

Support Vector Machines (SVM) classify data by identifying an optimal hyperplane that maximizes the margin between different classes. In an n-dimensional space, the hyperplane serves as a decision boundary that best separates data points into distinct categories. By maximizing the margin between the hyperplane and the nearest data points from each class, SVM ensures robustness and improved classification accuracy (Abe, 2005). The SVM model is implemented in GEE

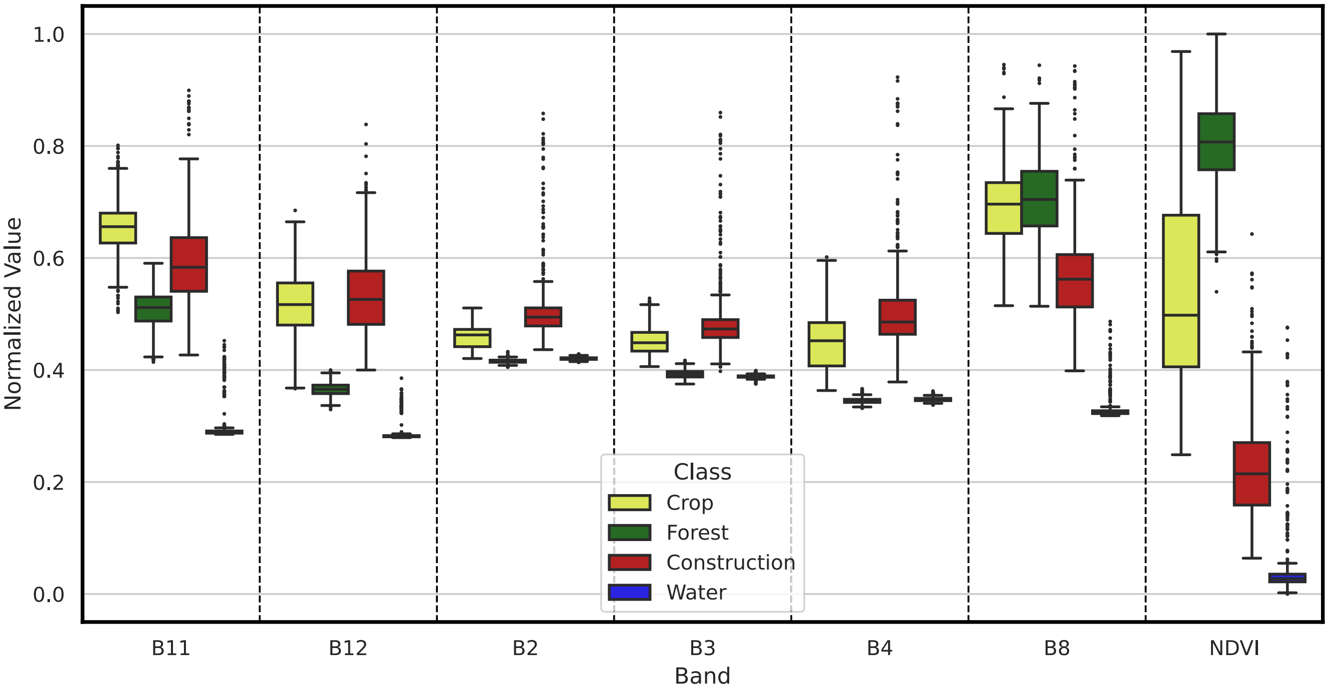


Fig. 2. Box-plots for band values by class (based on the training data).

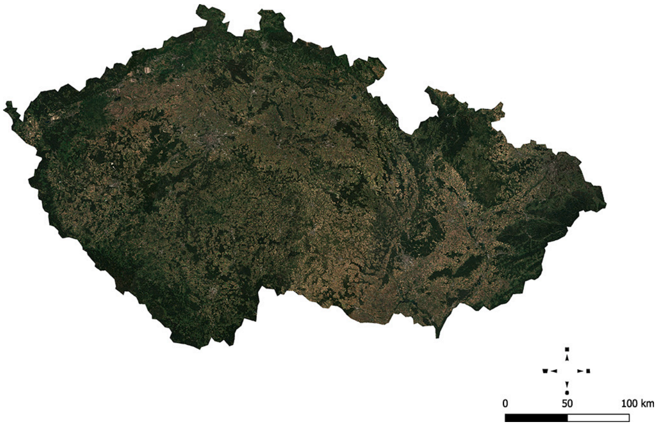


Fig. 3. Median cloud-masked Sentinel-2 image of the Czech Republic used for classification (January 1 to December 30, 2023).

using ee.Classifier.libsvm() with several tunable parameters, including decisionProcedure, which determines how the classifier finds the optimal hyperplane; svmType, which defines the type of SVM (e.g., C-SVC or one-class SVM); and kernelType, which specifies the function used to transform the feature space (linear, polynomial, or radial basis function). Additional parameters such as shrinking, degree (for polynomial kernels), gamma (kernel coefficient), coef0 (independent term in kernel function), cost (regularization parameter), nu (upper bound on the fraction of margin errors in nu-SVC), terminationEpsilon (stopping criterion tolerance), lossEpsilon, and oneClass can be fine-tuned using an iterative optimization process to maximize classification accuracy (Gorelick et al., 2017). In this study, we used the C-SVC type with a linear kernel while keeping the default values for the other parameters.

Naive Bayes is a probabilistic classification algorithm based on Bayes' Theorem, assuming independence among predictors. Despite its simplicity, Naive Bayes is highly efficient and effective, particularly in text classification tasks such as spam detection and sentiment analysis. It calculates the posterior probability of each class given a set of input features and assigns the observation to the class with the highest probability, leveraging frequency-based likelihood estimations from the training data (Aggarwal, 2014). The Naive Bayes classifier for remote sensing classifications is in GEE implemented using ee.Classifier.smileNaiveBayes() and was optimized by adjusting the lambda parameter, which controls the smoothing factor applied to prevent zero probabilities for unseen features. Different values of lambda were tested, and the one that achieved the highest classification accuracy while maintaining model stability was selected.

2.2.2. Accuracy assessment

To evaluate the performance of classification models in Google Earth Engine (GEE), several key metrics were utilized, including F-score, Kappa coefficient, Consumer Accuracy, Producer Accuracy, and Overall Accuracy. Each of these metrics provides unique insights into different aspects of model performance, allowing for a comprehensive evaluation of classification results. These metrics are derived from the confusion matrix, which records the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) (Cerulli, 2023).

Overall Accuracy represents the proportion of correctly classified samples relative to the total number of samples in the validation dataset. While this metric provides a general overview of model performance, it may not accurately reflect the model's effectiveness, particularly in cases of imbalanced data. In such scenarios, the accuracy metric can be misleading, as the model may perform well in terms of overall accuracy but poorly in correctly identifying underrepresented classes (Zhao et al., 2024). Therefore, other metrics like Kappa and F-score are often more informative in imbalanced classification tasks.

The Kappa Coefficient quantifies the agreement between predicted and actual classifications while accounting for the possibility of chance agreement. A Kappa value greater than 0 indicates better-than-random agreement, with 1 signifying perfect alignment. This metric is particularly useful for assessing classification reliability beyond random chance and offers a more nuanced understanding of model performance in comparison to Overall Accuracy (Zhao et al., 2024). In practical terms, the Kappa coefficient is useful for understanding the consistency of the model's predictions across all classes, particularly when the data is imbalanced.

Consumer Accuracy (also known as User Accuracy or Positive Predictive Value) measures the proportion of correctly classified samples within each predicted class. It indicates the probability that a sample assigned to a specific class truly belongs to that class, reflecting how reliable the model's predictions are for end users. A higher Consumer Accuracy means that users can trust the model's classification results for that class, particularly when making decisions based on the output (Zhao et al., 2024). This metric is particularly valuable in applications where the cost of misclassification (e.g., wrongly classifying a land cover type) is high.

Producer Accuracy (also referred to as Sensitivity) evaluates the proportion of correctly classified samples within each actual class. This metric provides insight into the model's ability to correctly identify samples of a given class, indicating the probability that an actual sample of a class will be correctly classified. High Producer Accuracy indicates that the model is effective in detecting all instances of a given class, even if those instances are fewer in number (Zhao et al., 2024). It is particularly important in situations where the model needs to avoid missing any instances of a class (e.g., detecting rare species or land cover types).

The F-score is the harmonic mean of precision and recall, offering a balanced metric for evaluating model performance, particularly in scenarios where there is a trade-off between precision and completeness. The F1 score ranges from 0 to 1, with 1 indicating perfect classification performance and 0 representing the lowest possible performance. This metric is especially valuable in cases involving imbalanced class distributions, as it provides a more balanced assessment than accuracy alone (Cerulli, 2023). By combining precision and recall, the F-score helps identify the model's ability to not only classify correctly but also ensure that it does so consistently across classes, especially in unbalanced datasets.

All of the above steps, including accessing satellite imagery, collecting training data, training the model, performing land cover classification, and calculating forest areas in individual regions, were carried out within the Google Earth Engine (GEE) environment. The workflow outlining these steps is illustrated in Fig. 4. This integrated approach within GEE allows for efficient processing of large-scale datasets and the application of machine learning models at a global scale, enhancing the accuracy and speed of environmental monitoring tasks such as land cover classification.

3. Results

The algorithms were trained using the classified training set, and their performance was evaluated on the test set. Table 2 presents the individual metrics for each model type. The Random Forest algorithm, which achieved the highest overall performance, was selected for the final prediction. This algorithm not only had the highest overall accuracy but also maintained a balanced accuracy across individual classes.

Random Forest outperformed the other models across nearly all metrics, achieving the highest overall accuracy of 98.3 %. The model demonstrated strong classification accuracy for all four classes. Its high Kappa value indicates excellent agreement and reliability, making it the most accurate model in this comparison.

The trained model is now capable of predicting new satellite images. Fig. 5 presents the classification of the Czech Republic using the estimated model. This classification was performed on the median Sentinel 2 image from January 1 to December 30, 2023, which is the same image

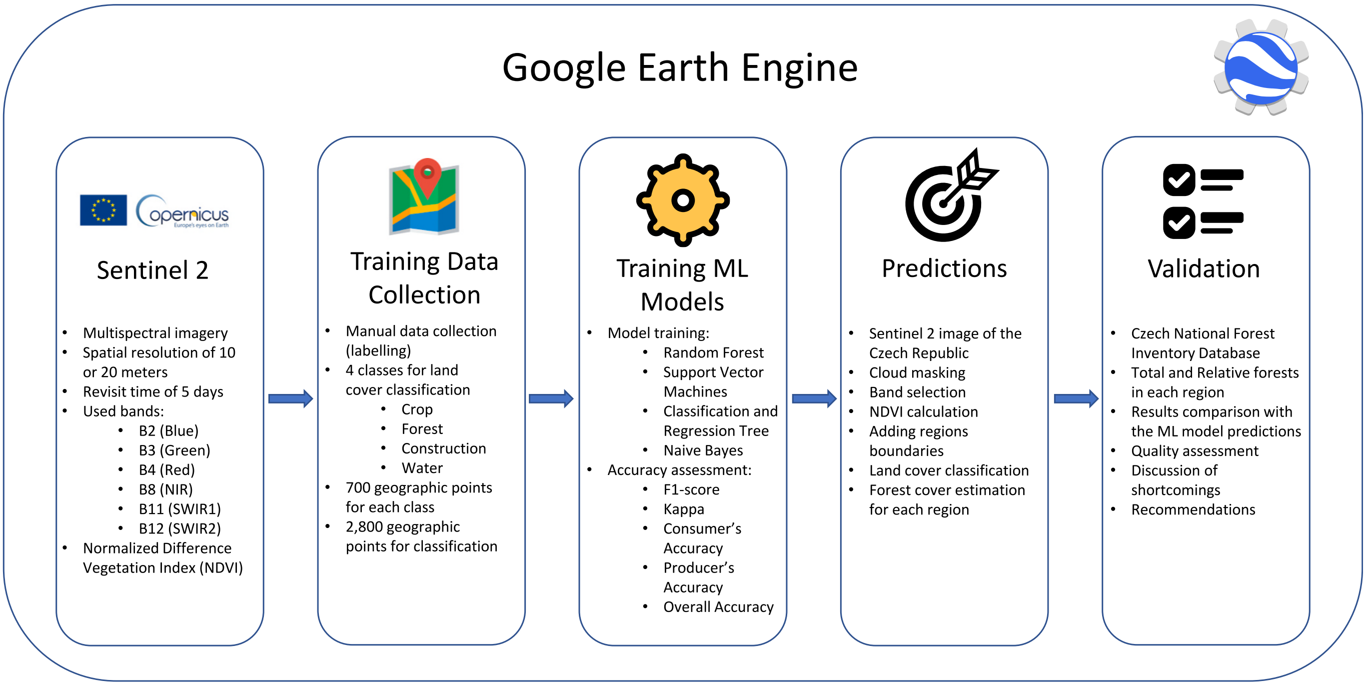


Fig. 4. Schematic representation of the methodological process.

Table 2
Comparison of accuracy metrics for classification algorithms.

Algorithm	Algorithm Setting	F1-score	Kappa	Consumer's Accuracy	Producer's Accuracy	Overall Accuracy
CART	Max. nodes: 13 Num. of trees: 1968	Crop: 0.931 Forest: 0.995 Construction: 0.939 Water: 1	0.956	Crop: 0.933 Forest: 0.995 Construction: 0.937 Water: 1	Crop: 0.928 Forest: 0.995 Construction: 0.941 Water: 1	0.968
Random Forest	Num. of trees: 100	Crop: 0.967 Forest: 0.997 Construction: 0.967 Water: 0.997	0.978	Crop: 0.955 Forest: 0.995 Construction: 0.980 Water: 1	Crop: 0.979 Forest: 1 Construction: 0.956 Water: 0.995	0.983
Support Vector Machines	Type: C-SVC Kernel: Linear	Crop: 0.911 Forest: 0.989 Construction: 0.926 Water: 0.995	0.942	Crop: 0.935 Forest: 0.977 Construction: 0.910 Water: 1	Crop: 0.887 Forest: 1 Construction: 0.941 Water: 0.991	0.957
Naive Bayes	Lambda: 0.001	Crop: 0.785 Forest: 0.928 Construction: 0.831 Water: 0.932	0.828	Crop: 0.883 Forest: 0.866 Construction: 0.782 Water: 0.966	Crop: 0.701 Forest: 1 Construction: 0.888 Water: 0.900	0.871

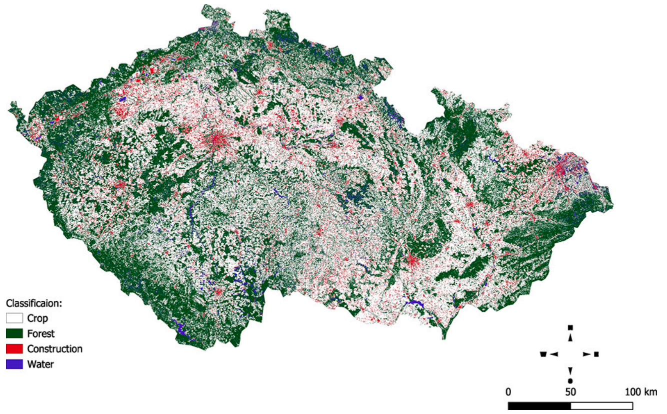


Fig. 5. Classification results of the Czech Republic satellite image using the Random Forest model.

shown in Fig. 3. The median image is a composite of multiple images collected within this specified date range.

The developed model was further used to compare its results with those of the National Forest Inventory (NFI), specifically with the most recent inventory conducted between 2016 and 2020, and evaluated until 2023. For this comparison, the afforestation status in individual regions of the Czech Republic was estimated, and the relative occurrence of forests in these regions was calculated. The classification was based on the median image from the period of June 6 to October 31, 2018. A summary of the share of each classified class in the respective regions of the Czech Republic is provided in Table 3.

These results were then compared with published NFI estimates (Institute for Forest Management, 2024). A comparison of the model's estimated results with the NFI estimates is shown in Table 4. The comparison includes estimates of forest area in individual regions and their relative representation. These results were contrasted with the findings from the latest NFI survey, which indicates the range of forest areas and their relative proportions across different regions.

Table 3
Estimated distribution of classified classes in regions of the Czech Republic (2018, thousand ha).

Region	Crop	Forest	Construction	Water
Central Bohemian Region	670.00	292.30	104.62	25.26
South Bohemian Region	509.22	420.15	295.00	50.21
Plzeň Region	373.72	321.70	48.86	20.85
Karlovy Vary Region	137.87	159.24	16.27	17.29
Ustí nad Labem Region	268.30	172.33	76.74	14.73
Liberec Region	139.46	143.19	16.53	17.41
Hradec Králové Region	259.48	119.83	56.83	39.18
Pardubice Region	259.64	114.38	45.26	32.82
Vysočina Region	423.35	185.76	27.63	42.25
South Moravian Region	437.85	191.11	75.28	14.22
Olomouc Region	290.01	165.91	45.58	26.54
Zlín Region	172.25	184.17	29.54	9.37
Moravian-Silesian Region	267.30	204.77	45.26	25.01
Prague	24.78	6.63	17.49	1.13

Table 4
Comparison of Random Forest model estimations with NFI results: total forest area in thousands of ha and forest share in % in regions of the Czech Republic.

Region	MODEL (thousand *ha*)	MODEL (%)	NFI (thousand *ha*)	NFI (%)
Central Bohemian Region	292.30	26.76	321.2–352.8	29.5–31.9
South Bohemian Region	420.15	32.96	398.3–433.1	40.0–42.8
Plzeň Region	321.70	42.05	321.6–353.2	42.4–45.6
Karlovy Vary Region	159.24	48.16	155.5–177.9	47.8–52.6
Ustí nad Labem Region	172.33	32.39	184.4–208.6	35.1–38.7
Liberec Region	143.19	45.23	147.6–169.4	47.6–52.6
Hradec Králové Region	119.83	25.21	146.2–168.0	31.3–35.1
Pardubice Region	114.38	25.30	134.6–155.6	30.2–34.0
Vysočina Region	185.76	27.36	202.8–228.2	30.2–33.4
South Moravian Region	191.11	26.60	201.8–227.2	28.3–31.3
Olomouc Region	165.91	31.42	182.8–207.0	35.2–38.8
Zlín Region	184.17	46.59	160.5–183.3	41.4–45.8
Moravian-Silesian Region	204.77	37.76	192.5–217.3	35.9–39.5
Prague	6.63	13.25	4.8–9.6	10.1–18.9

The model successfully identified areas of afforestation and the relative representation of forests, with results that in many cases fall within the NFI's estimated ranges. Specifically, for five regions, the forest area values were within the NFI interval, and in other cases, the values were very close to the NFI's reported ranges.

The largest deviation from the interval limits occurred in the Central Bohemia region, with a value of 28.9 thousand hectares. The model's forest area estimates fell within the NFI interval for a total of five regions. The relative representation of forest area in the regions matched the NFI intervals in three cases. The largest deviation in relative representation was observed in the South Bohemian Region, with a difference of 7.04 %. In most cases, the model tended to underestimate the NFI statistics.

Finally, the total forest area in the Czech Republic was calculated. According to the Random Forest model, the forest area is 2681.47 thousand hectares. The Food and Agriculture Organization of the United Nations (2025) estimates this value at approximately 2677.09 thousand hectares. The official NFI (Institute for Forest Management, 2024) results report a forest area of 2923.2 ± 37.6 thousand hectares.

4. Discussion

Validating machine learning models for forest management is essential to ensure their reliability and enable informed decision-making by forest managers (Janová et al., 2024). This study aimed to assess the capability of a machine learning model, trained on Sentinel-2 satellite imagery, to produce forest cover estimates comparable to those provided by the National Forest Inventory (NFI). Using Google Earth Engine (GEE), we implemented a classification task involving four target classes, with the Random Forest algorithm emerging as the best-performing model due to its high accuracy and robustness. The superiority of Random Forest for land cover classification tasks has been widely documented (Becker et al., 2021), (Bajaj et al., 2024), and its flexibility extends beyond classification to regression tasks (Choudhary et al., 2022). While simpler models such as Naive Bayes offer computational efficiency, they often sacrifice accuracy (Pan et al., 2022).

Random Forest is a robust ensemble learning technique that typically excels in handling high-dimensional data, as it creates multiple decision trees by bootstrapping subsets of the data and averaging their predictions. This ensemble approach helps mitigate the risk of overfitting, particularly when working with noisy or complex data, as is often the case in remote sensing applications (Pal, 2005). Random Forest also has a built-in feature selection mechanism, which helps in identifying the most relevant variables and reduces the impact of irrelevant or redundant features (Cutler et al., 2007). Additionally, it is less sensitive to hyperparameter tuning, which can often be a challenge for other models such as Support Vector Machines (SVM) or Naive Bayes. These strengths allowed RF to perform consistently well in our study, as it effectively handled the multicollinearity and large number of input features typical of remote sensing data, leading to higher classification accuracy. In comparison, other models such as CART, SVM, and Naive Bayes have specific limitations that likely contributed to their slightly lower performance. CART, while interpretable and straightforward, tends to overfit, particularly when the tree is not properly pruned, and struggles with high-dimensional data, making it less effective in our context. SVM, on the other hand, is highly effective in scenarios with clear class boundaries but requires careful selection of kernel functions and tuning of parameters such as the regularization term, which can affect its performance on noisy or imbalanced datasets (Mountrakis et al., 2011). Finally, Naïve Bayes assumes independence between features, a strong assumption that is often violated in real-world datasets like ours, where vegetation indices or individual bands can be highly correlated, resulting in suboptimal performance (Pedrayes et al., 2023).

The Random Forest model demonstrated reliable forest classification across most regions. The classification results indicated that the model was particularly effective in regions such as Karlovy Vary, Plzeň, Moravian-Silesian, and Prague, where estimates of afforestation fell within the NFI's reported intervals. This suggests a high degree of reliability for these areas. However, the model demonstrated notable discrepancies in Central Bohemia and Hradec Králové, where forest area estimates deviated significantly from NFI values. The largest deviation was observed in Central Bohemia, with a difference of 28.9 thousand hectares, while relative representation in the South Bohemian Region diverged by 7.04 % from the NFI's estimates. Notably, the total forest area estimated by the model (2681.47 thousand hectares) closely aligns with FAO's estimate (2677.09 thousand hectares), further validating the model's applicability. However, discrepancies with the official NFI estimates (2923.2 ± 37.6 thousand hectares) indicate the necessity for further refinement.

A notable pattern observed across several regions was the model's tendency to underestimate forest area and relative representation. This underestimation aligns with findings from prior research, where satellite-based models often yield conservative estimates due to classification errors, mixed pixels, or insufficient training data (Feizizadeh et al., 2023). Such discrepancies highlight the importance of complementary validation methods, including on-the-ground verification and alternative remote sensing techniques.

Our study's reliance on Sentinel-2 data aligns with the methodologies employed by similar studies (Tamiminia et al., 2024) but also reveals potential areas for refinement. For instance, Zurqani (2025) demonstrated that integrating multiple data sources, including LiDAR, significantly enhances model accuracy when estimating forest biomass. The ability of LiDAR to capture vertical forest structures makes it particularly valuable for biomass estimation, a factor that satellite imagery alone cannot fully address (Zhang et al., 2025). The integration of LiDAR with Sentinel-2 data could therefore enhance classification accuracy and compensate for some of the observed underestimations.

Building on this, future research could explore the fusion of Sentinel-2 imagery with other remote sensing data sources, such as LiDAR or Synthetic Aperture Radar (SAR), to improve model performance and reliability. Combining the spectral information from Sentinel-2 with the structural insights provided by LiDAR or the all-weather capabilities of SAR could offer a more comprehensive view of forest characteristics (Zhang et al., 2025). This multi-sensor approach has the potential to enhance not only biomass estimation but also other forest parameters such as species diversity and forest health.

Additionally, Kalinaki et al. (2023) showcased the effectiveness of deep learning models in analysing vegetation dynamics, particularly in rapidly changing landscapes. Their study in Brunei revealed that deep learning techniques, such as convolutional neural networks (CNNs), outperformed traditional machine learning algorithms in identifying spatial-temporal patterns in land cover changes. Given our model's limitations in certain regions, incorporating deep learning approaches could improve accuracy by allowing the model to better recognize complex spatial patterns and spectral variations (Cao et al., 2021), (Zhao et al., 2024). This is particularly relevant for heterogeneous landscapes where traditional classifiers may struggle.

A major challenge in applying machine learning models for forest monitoring lies in ensuring that predictions remain reliable across different ecological contexts. While our model demonstrated strong performance in certain regions, its accuracy varied when applied to diverse landscapes. Similar studies have noted that even highly accurate models can struggle when applied to new geographic regions without sufficient recalibration (Feizizadeh et al., 2023). To address this challenge, continuous model validation using new data sources is crucial. Periodic recalibration of machine learning models with updated training data can help mitigate shifts in land cover patterns and improve long-term applicability.

The integration of multi-source remote sensing data could also enhance model robustness. Lechner et al. (2020) emphasized the value of combining Sentinel-2 with Synthetic Aperture Radar (SAR) and high-resolution aerial imagery to improve classification accuracy. SAR, in particular, can provide valuable data in areas affected by cloud cover, which often poses challenges for optical satellite sensors like Sentinel-2. Combining these data sources could address some of the limitations observed in our study, particularly in regions where Sentinel-2 alone was insufficient for accurate classification.

Another key challenge is the availability and quality of training datasets (Janová and Hampel, 2016). Machine learning models require large, diverse, and high-quality ground-truth data to generalize effectively. However, obtaining verified training data, particularly for species composition and timber volume estimation, remains an obstacle. The reliance on NFI data provides a benchmark for validation, but further efforts are needed to improve the diversity of training datasets. The work of Pan et al. (2022) highlighted the importance of integrating field surveys with remotely sensed data to refine classification models and reduce bias. Addressing these data limitations is essential for enhancing the reliability of machine learning models in forest monitoring.

While machine learning models, including Random Forest, offer valuable insights for forest management, their application must be accompanied by careful validation and consideration of data limitations. The integration of multiple data sources and advanced machine learning techniques, such as deep learning, holds the potential to improve model accuracy and applicability. However, to ensure their effectiveness in real-world applications, continued efforts in model validation, the development of high-quality training datasets, and the integration of complementary data types will be necessary (Fassnacht et al., 2024).

5. Conclusion

This paper explored the potential of using machine learning techniques applied to Sentinel-2 satellite data to optimize certain processes within the National Forest Inventory (NFI). The results demonstrate that satellite imagery can provide reliable estimates for tasks such as estimating forest area and its relative distribution across regions. Our approach presents a promising, cost-effective alternative to traditional NFI methods, offering a viable solution for streamlining forest inventory processes. Considering the substantial resources required by conventional NFI approaches, this study advocates for further investigation into the integration of satellite data and other advanced technologies to improve forest monitoring, enhance data accuracy, and reduce the financial burden of data collection.

CRediT authorship contribution statement

Richard Kovárník: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Jitka Janová:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Supported by the grant IGA25-PEF-DP-004 of the IGA PEF MENDELU Grant Agency.

Data availability

The code and data used in this study are available at GitHub Repository: https://github.com/xkovarn1/Sentinel-2-NFI-validation

References

Abe, S., 2005. Support Vector Machines for Pattern Classification. Springer, London.

Aggarwal, C.C., 2014. Data Classification: Algorithms and Applications. CRC Press.

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., Moghaddam, S.H., Mahdavi, S., Ghahremanloo, M., Parsian, S., et al., 2020. Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. 13, 5326–5350. https://doi.org/10.1109/JSTARS.2020.3021052.

Bajaj, M., Sasaki, N., Tsusaka, T.W., Venkatappa, M., Abe, I., Shrestha, R.P., 2024. Assessing changes in mangrove forest cover and carbon stocks in the lower Mekong region using Google earth engine. Innov. Green Developm. 3, 100140. https://doi.org/10.1016/j.igd.2024.100140.

Becker, W.R., Ló, T.B., Johann, J.A., Mercante, E., 2021. Statistical features for land use and land cover classification in Google earth engine. Remote Sens. Appl. Soc. Environ. 21, 100459. https://doi.org/10.1016/j.rsase.2020.100459.

Bouckaert, R.R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., Scuse, D., 2016. WEKA Manual for Version 3-9-1. Hamilton.

Bountos, N.I., Ouaknine, A., Rolnick, D., 2023. FoMo-bench: a multi-modal, multi-scale and multi-task Forest monitoring benchmark for remote sensing foundation models. arXiv. https://doi.org/10.48550/arXiv.2312.10114 preprint arXiv:2312.10114.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1007/BF00058655.

Cao, J., Zhang, Z., Luo, Y., Zhang, L., Zhang, J., Li, Z., Tao, F., 2021. Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. Eur. J. Agron. 123, 126204. https://doi.org/10.1016/j.eja.2020.126204.

Cerulli, G., 2023. Fundamentals of Supervised Machine Learning: With Applications in Python, R, and Stata. Springer International Publishing.

Chen, S., Woodcock, C.E., Bullock, E.L., Arévalo, P., Torchinava, P., Peng, S., Olofsson, P., 2021. Monitoring temperate forest degradation on Google Earth Engine using Landsat time series analysis. Remote Sens. Environ. 265, 112648. https://doi.org/10.1016/j.rse.2021.112648.

Choudhary, K., Shi, W., Dong, Y., Paringer, R., 2022. Random Forest for rice yield mapping and prediction using Sentinel-2 data with Google Earth Engine. Adv. Space Res. 70, 2443–2457. https://doi.org/10.1016/j.asr.2022.06.073.

Coops, N.C., Tompalski, P., Goodbody, T.R., Achim, A., Mulverhill, C., 2023. Framework for near real-time forest inventory using multi source remote sensing data. Forestry 96, 1–19. https://doi.org/10.1093/forestry/cpac015.

Cutler, D.R., Edwards Jr., T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88 (11), 2783–2792. https://doi.org/10.1890/07-0539.1.

European Commission, 2020. Horizon 2020. Horizon 2020.

Fassnacht, F.E., White, J.C., Wulder, M.A., Næsset, E., 2024. Remote sensing in forestry: current challenges, considerations and directions. Forestry Int. J. For. Res. 97, 11–37. https://doi.org/10.1093/forestry/cpad024.

Feizizadeh, B., Omarzadeh, D., Kazemi Garajeh, M., Lakes, T., Blaschke, T., 2023. Machine learning data-driven approaches for land use/cover mapping and trend analysis using Google Earth Engine. J. Environ. Plan. Manag. 66, 665–697. https://doi.org/10.1080/09640568.2021.2001317.

Food and Agriculture Organization of the United Nations, 2025. Global Forest Resources Assessment. https://fra-data.fao.org/assessments/fra/2020/CZE/home/overview.

Genuer, R., Poggi, J.M., 2020. Random Forests with R. Springer International Publishing.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.06.031.

Gumma, M.K., Takashi, Y., Panjala, P., Deevi, K.C., Inthavong, V., Bellam, P.K., Mohammed, I., 2023. Assessment of cropland changes due to new canals in Vientiane prefecture of Laos using earth observation data. Smart Agric. Technol. 4, 100149. https://doi.org/10.1016/j.atech.2022.100149.

Huang, S., Tang, L., Hupy, J.P., Wang, Y., Shao, G., 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J. For. Res. 32, 1–6. https://doi.org/10.1007/s11676-020-01155-1.

Institute for Forest Management, 2024. NIL - National Forest Inventory. NIL - National Forest Inventory.

Janová, J., Hampel, D., 2016. Optimal managing of forest structure using data simulated optimal control. CEJOR 24 (2), 297–307. https://doi.org/10.1007/s10100-015-0383-x.

Janová, J., Bödeker, K., Bingham, L., Kindu, M., Knoke, T., 2024. The role of validation in optimization models for forest management. Ann. For. Sci. 81, 19. https://doi.org/10.1186/s13595-024-01235-w.

Kalinaki, K., Malik, O.A., Lai, D.T., Sukri, R.S., Wahab, R.B., 2023. Spatial-temporal mapping of forest vegetation cover changes along highways in Brunei using deep learning techniques and Sentinel-2 images. Eco. Inform. 77, 102193. https://doi.org/10.1016/j.ecoinf.2023.102193.

Kangas, A., Maltamo, M., 2006. Forest Inventory: Methodology and Applications, Vol. 10. Springer Science & Business Media.

Koskinen, J., Leinonen, U., Vollrath, A., Ortmann, A., Lindquist, E., d'Annunzio, R., Pekkarinen, A., Käyhkö, N., 2019. Participatory mapping of forest plantations with open Foris and Google Earth Engine. ISPRS J. Photogramm. Remote Sens. 148, 63–74. https://doi.org/10.1016/j.isprsjprs.2018.12.011.

Lechner, A.M., Foody, G.M., Boyd, D.S., 2020. Applications in remote sensing to forest ecology and management. One Earth 2, 405–412. https://doi.org/10.1016/j.oneear.2020.05.001.

McRoberts, R.E., Tomppo, E.O., 2007. Remote sensing support for national forest inventories. Remote Sens. Environ. 110, 412–419. https://doi.org/10.1016/j.rse.2006.09.034.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66 (3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001.

Pal, M., 2005. Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26 (1), 217–222. https://doi.org/10.1080/01431160412331269698.

Pan, X., Wang, Z., Gao, Y., Dang, X., Han, Y., 2022. Detailed and automated classification of land use/land cover using machine learning algorithms in Google earth engine. Geocarto Int. 37, 5415–5432. https://doi.org/10.1080/10106049.2021.1917005.

Panahi, H., Azizi, Z., Kiadaliri, H., Almodaresi, S.A., Aghamohamadi, H., 2024. Bare soil detecting algorithms in western Iran woodlands using remote sensing. Smart Agric. Technol. 7, 100429. https://doi.org/10.1016/j.atech.2024.100429.

Pedrayes, O.D., Usamentiaga, R., Trichakis, Y., Bouraoui, F., 2023. Remote sensing for detecting freshly manured fields. Eco. Inform. 75, 102006. https://doi.org/10.1016/j.ecoinf.2023.102006.

Pérez-Cutillas, P., Pérez-Navarro, A., Conesa-Garcia, C., Zema, D.A., Amado-Alvarez, J. P., 2023. What is going on within google earth engine? A systematic review and meta-analysis. Remote Sens. Appl. Soc. Environ. 29, 100907. https://doi.org/10.1016/j.rsase.2022.100907.

Pinna, D., Pezzuolo, A., Cogato, A., Pornaro, C., Macolino, S., Marinello, F., 2024. Applications of satellite platforms and machine learning for mapping and monitoring grasslands and pastures: A systematic and comprehensive review. Smart Agric. Technol. 100571. https://doi.org/10.1016/j.atech.2024.100571.

Savitha, C., Talari, R., 2023. Mapping cropland extent using sentinel-2 datasets and machine learning algorithms for an agriculture watershed. Smart Agric. Technol. 4, 100193. https://doi.org/10.1016/j.atech.2023.100193.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google earth engine for geo-big data applications: A meta-analysis and systematic review. ISPRS J. Photogramm. Remote Sens. 164, 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001.

Tamiminia, H., Salehi, B., Mahdianpari, M., Goulden, T., 2024. State-wide forest canopy height and aboveground biomass map for New York with 10 m resolution, integrating GEDI, Sentinel-1, and Sentinel-2 data. Eco. Inform. 102404. https://doi.org/10.1016/j.ecoinf.2023.102404.

Terribile, F., Acutis, M., Agrillo, A., Anzalone, E., Azam-Ali, S., Bancheri, M., Baumann, P., Birli, B., Bonfante, A., Botta, M., et al., 2024. The LANDSUPPORT geospatial decision support system (S-DSS) vision: operational tools to implement sustainability policies in land planning and management. Land Degrad. Dev. 35, 813–834. https://doi.org/10.1002/ldr.4954.

Zhang, B., Wang, Z., Ma, T., Wang, Z., Li, H., Ji, W., He, M., Jiao, A., Feng, Z., 2025. Correcting forest aboveground biomass biases by incorporating independent canopy height retrieval with conventional machine learning models using GEDI and ICESat-2 data. Eco. Inform. 103045. https://doi.org/10.1016/j.ecoinf.2025.103045.

Zhao, Z., Islam, F., Waseem, L.A., Tariq, A., Nawaz, M., Islam, I.U., Bibi, T., Rehman, N. U., Ahmad, W., Aslam, R.W., et al., 2024. Comparison of three machine learning algorithms using Google earth engine for land use land cover classification. Rangel. Ecol. Manag. 92, 129–137. https://doi.org/10.1016/j.rama.2023.10.007.

Zurqani, H.A., 2025. A multi-source approach combining GEDI LiDAR, satellite data, and machine learning algorithms for estimating forest aboveground biomass on Google earth engine platform. Eco. Inform. 103052. https://doi.org/10.1016/j.ecoinf.2025.103052.

Zurqani, H.A., Post, C.J., Mikhailova, E.A., Allen, J.S., 2019. Mapping urbanization trends in a forested landscape using Google Earth Engine. Remote Sens. Earth Syst. Sci. 2, 173–182. https://doi.org/10.1007/s41976-019-00020-y.