


## Article

# Improving Forest Detection Using Machine Learning and Remote Sensing: A Case Study in Southeastern Serbia

Ivan Potić<sup>1,†</sup>, Zoran Srdić<sup>1,†</sup>, Boris Vakanjac<sup>1</sup>, Saša Bakrač<sup>1,2,\*</sup>, Dejan Đorđević<sup>1,2</sup>, Radoje Banković<sup>1,2</sup> and Jasmina M. Jovanović<sup>3</sup>

<sup>1</sup> Military Geographical Institute “General Stevan Bošković”, 11000 Belgrade, Serbia; ipotic@gmail.com (I.P.); zoran.m.srdic@gmail.com (Z.S.); borivac@gmail.com (B.V.); dejan.r.djordjevic@vs.rs (D.Đ.); radoje.bankovic@vs.rs (R.B.)

<sup>2</sup> Military Academy, University of Defense, 11000 Belgrade, Serbia

<sup>3</sup> Faculty of Geography, University of Belgrade, 11000 Belgrade, Serbia; jasmina.jovanovic@gef.bg.ac.rs

\* Correspondence: sasa.bakrac@vs.rs; Tel.: +381-113205009

† Co-first authors; these authors contributed equally to this work.

**Featured Application:** The primary application of this work is in environmental resource management, specifically in the detection and monitoring of vegetation patterns and changes. By employing a machine learning approach, specifically the Support Vector Machines (SVM) algorithm, the study demonstrates that including vegetation indices alongside multispectral bands significantly improves the accuracy of vegetation detection, achieving an overall classification accuracy of up to 99.01%. The study’s findings underscore the potential of machine learning and remote sensing in vegetation detection and monitoring and highlight the importance of incorporating vegetation indices to enhance classification accuracy. The matter above has significant implications for decision-making processes in environmental resource management, particularly in regions with diverse forest ecosystems. The potential applications of this work extend beyond the specific geographical context of the study. The methodology and findings could be applied to other regions and ecosystems, providing valuable insights for the preservation and conservation of forest ecosystems globally. Future research could further explore the applicability of these findings in different geographical regions and investigate other vegetation indices to improve the accuracy of forest detection and monitoring processes.

**Abstract:** Vegetation plays an active role in ecosystem dynamics, and monitoring its patterns and changes is vital for effective environmental resource management. This study explores the possibility of machine learning techniques and remote sensing data to improve the accuracy of forest detection. The research focuses on the southeastern part of the Republic of Serbia as a case study area, using Sentinel-2 multispectral bands. The study employs publicly accessible satellite data and incorporates different vegetation indices to improve classification accuracy. The main objective is to examine the practicability of expanding the input parameters for forest detection using a machine learning approach. The classification process is performed by employing support vector machines (SVM) algorithm and utilising the SVM module in the scikit-learn package. The results demonstrate that including vegetation indices alongside the multispectral bands significantly improves the accuracy of vegetation detection. A comprehensive assessment reveals an overall classification accuracy of up to 99.01% when the selected vegetation indices (MCARI, RENDVI, NDI45, GNDVI, NDII) are combined with the Sentinel-2 bands. This research highlights the potential of machine learning and remote sensing in forest detection and monitoring. The findings underscore the importance of incorporating vegetation indices to enhance classification accuracy using the Python programming language. The study’s outcomes provide valuable insights for environmental resource management and decision-making processes, particularly in regions with diverse forest ecosystems.



**Citation:** Potić, I.; Srdić, Z.; Vakanjac, B.; Bakrač, S.; Đorđević, D.; Banković, R.; Jovanović, J.M. Improving Forest Detection Using Machine Learning and Remote Sensing: A Case Study in Southeastern Serbia. *Appl. Sci.* **2023**, *13*, 8289. <https://doi.org/10.3390/app13148289>

Academic Editors: Romano Lottering, Kabir Peerbhay and Samuel Adelabu

Received: 12 June 2023

Revised: 10 July 2023

Accepted: 10 July 2023

Published: 18 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** vegetation detection; remote sensing; Python; machine learning; classification accuracy; Sentinel-2

---

## 1. Introduction

Vegetation is an essential component of ecosystems that connects the atmospheric, hydrological, and pedological processes [1]. Environmental preservation and conservation heavily depend on economic stability and human and political resources. In the past two decades, the cost of evaluation procedures has significantly reduced due to the public availability of open satellite data. Now, we can obtain crucial information on deforestation and degradation by employing remote sensing techniques to analyse this data [2–5]. Earth observation (EO) data, which include satellite, aerial, or ground-based observations, and geospatial data are crucial for monitoring changes in forest ecosystems, especially for identifying vegetation degradation [6,7]. Monitoring land use and land cover change is vital to the ecosystem. Remote sensing offers excellent potential for monitoring landscape change caused by natural cycles and human activity [8]. One of the crucial applications of remote sensing in environmental resource management and decision-making is the detection and quantitative evaluation of vegetation patterns. This technology is pivotal in assessing the ecosystem and identifying vegetation patterns and structural shifts. Such assessments and identifications are principal when evaluating and monitoring natural resources. Remote sensing in forest detection has been a significant research and development topic. Remote sensing technologies provide a powerful tool for monitoring and managing forests on a large scale, offering the ability to detect changes in forest cover and health over time. Barmpoutis et al. (2020) provide an overview of optical remote sensing technologies used in early fire warning systems, highlighting the importance of these technologies in mitigating the impacts of natural hazards such as large-scale forest fires [9]. Similarly, Housman et al. (2018) discuss the Operational Remote Sensing (ORS) program, which leverages Landsat and MODIS data to detect forest disturbances across the United States. The ORS program supplements traditional Insect and Disease Survey (IDS) data with imagery-derived forest disturbance data, demonstrating the potential of remote sensing in forest health monitoring [10]. Furthermore, Chen et al. (2018) present a novel approach to individual tree-level forest inventory using airborne LiDAR (Light Detection And Ranging) remote sensing. Their research underscores the potential of remote sensing technologies in providing detailed, high-resolution data for forest management [11].

This study's primary challenge is identifying forest cover. In this case, forest detection is simplified to a classification issue involving categorising the input data set into two classes, "forest" and "not forest". This binary classification is suitable for inventorying forests or creating thematic masks for topographic maps when it is essential to identify forest cover without distinguishing forest types. In terms of binary classification, this approach offers several advantages. Binary classification simplifies the problem by focusing on the distinction between two classes, which can lead to more accurate and efficient models. It is beneficial when classes are imbalanced, allowing the model to focus on the minority class. Furthermore, binary classification models are often easier to interpret and understand, making them more practical for decision-making processes [12].

Forest detection becomes challenging and complex in such instances, especially in regions with high biodiversity, i.e., a wide range of forest ecosystems. The challenge of categorising data under such specific conditions is distinctively novel. The research investigated the feasibility of augmenting the initial array of input variables, including Sentinel-2 bands and vegetation indices, for executing the machine learning protocol. Satellite imagery serves as the input data for forest identification and is the most prevalent data source for forest inventory, particularly for categorising extensive regions. Examining the content within satellite imagery presents an additional concern, primarily due to the heterogeneity of materials in the images and the substantial volume of data. It is imperative

to employ sophisticated and robust technologies to effectively manage such a complex data set, especially in categorisation tasks. In addressing the classification problem associated with forest detection using satellite imagery, this research harnessed the power of artificial intelligence and machine learning.

The development of Machine Learning (ML) requires the determination of all essential metrics for the decision-making process. The mechanisms of machine learning generate models to enhance the metrics. In order to ensure the development of an effective solution for any decision-making process, it is crucial to carefully select and consider the metrics used throughout the conceptual phases. This is necessary because the metrics are essential in decision-making, and their selection can significantly impact the outcome [13]. ML's purpose is to anticipate future occurrences or situations unknown to the computer. It belongs to the subfield of artificial intelligence (AI) that synthesises the underlying correlations between data and information via the systematic application of algorithms [14]. In 1959, Arthur Samuel defined ML as "the field of study that allows computers to learn without being explicitly programmed". He stated that training computers to learn from experience would someday obviate the need for a significant portion of this comprehensive programming work [15]. The increasing prevalence of ML can be attributed to its ability to describe underlying connections within massive data arrays, thereby solving challenges in big data analytics, behavioural pattern identification, and information evolution. In addition, ML systems may be taught to classify the changing circumstances of a process to represent changes in operational behaviour. As knowledge evolves under the impact of new ideas and technologies, ML systems may detect disruptions to old models and redesign and retrain themselves to adapt to and coevolve with the new information [16,17].

Using vegetation indices and multispectral bands in machine learning models has proven to be a powerful tool in various applications. For instance, researchers have successfully used vegetation indices derived from light reflectance properties of plants to distinguish soybean from weeds, demonstrating the potential of these indices as decision-support tools for weed identification [18]. In precision agriculture, an automatic segmentation method combining vegetation indices with a Discriminative Common Vector Approach classification algorithm has outperformed traditional methods, facilitating sustainable production [19]. Furthermore, a machine learning model using the extreme gradient boosting method is developed to predict vegetation growth throughout the growing season in China, highlighting the potential of these techniques for monitoring vegetation dynamics and crop growth [20]. Lastly, the selection of suitable Sentinel-2 bands and vegetation index for crop classification using artificial neural networks has been discussed, underscoring the importance of these parameters in enhancing classification accuracy [21]. Mentioned articles demonstrate the potential of incorporating vegetation indices and multispectral channels in machine learning models for improved vegetation detection.

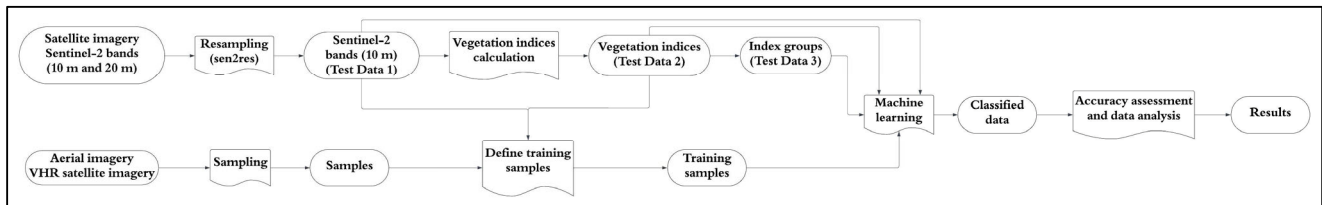
This study aims to leverage the power of remote sensing technologies, specifically focusing on the Support Vector Machines (SVM) algorithm for forest detection and classification. The primary objectives of the research are:

- To explore the potential of remote sensing in detecting and classifying forests, with a particular focus on binary classification;
- To define optimal parameters of the SVM algorithm, specifically the C and gamma parameters, for effective forest classification;
- To evaluate the advantages of binary classification in forest detection and discuss its implications for environmental management and conservation;
- To contribute to the existing body of knowledge by introducing an original approach to forest detection using remote sensing technologies.

The study's findings are expected to provide valuable insights into the application of remote sensing and SVM in forest detection, potentially informing future research and practices in the field.

## 2. Materials and Methods

The method used the SVM algorithm for satellite imagery classification. In addition to selected Sentinel-2 multispectral bands, vegetation indices were added to study their ability to increase classification accuracy (Figure 1) individually and as a group of indices.

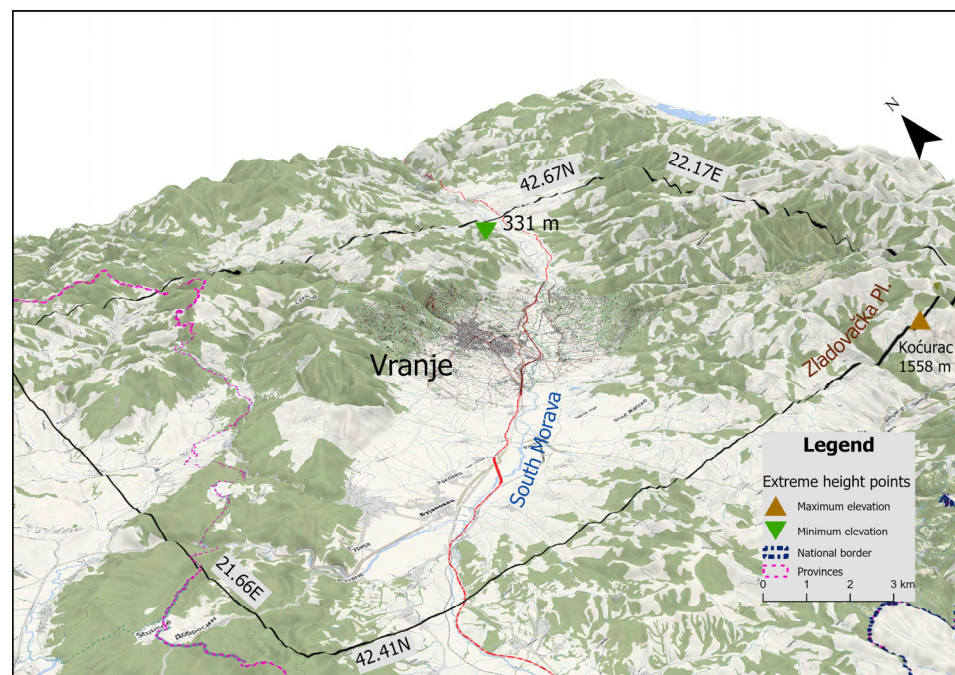


**Figure 1.** Improving Forest Detection Using Machine Learning and Remote Sensing workflow chart.

### 2.1. Study Area

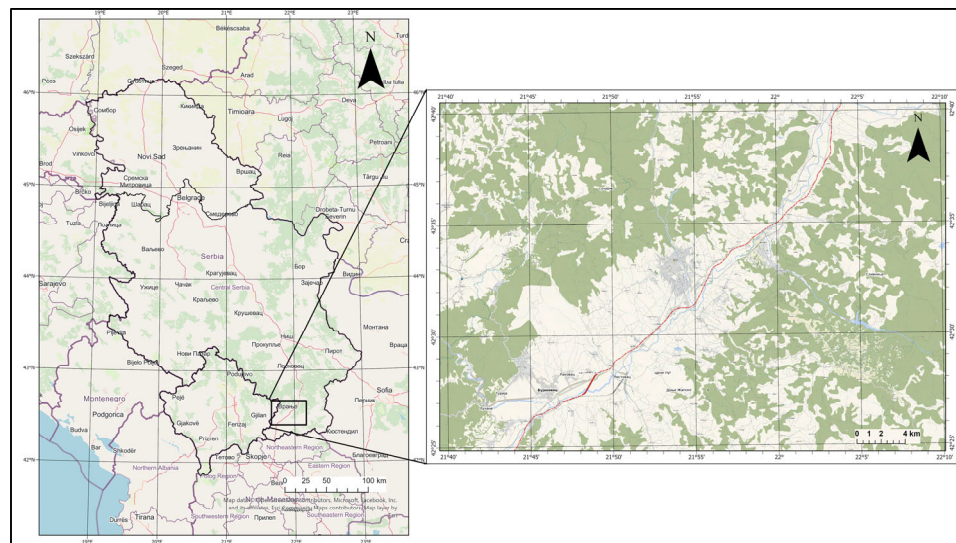
The southeastern part of the Republic of Serbia was chosen as the area of interest (AOI) (Figure 3). The area covered 1218 km<sup>2</sup> (42 × 29 km) with the central point at 575,100 and 4,710,500 (34T UTM/WGS84) or 21.9147 E, 42.5429 N (WGS84) coordinates.

The city of Vranje is located in the centre of the study area. In a southwest-to-northeast direction, the region is intersected by the South Morava River, which creates a vast, flat region. The region's northwest consists of hilly terrain, whereas the southeast is dominated by mountainous terrain. The lowest point is situated in the valley of the South Morava River, on the northern boundary 331 m above sea level (a.s.l.). The highest point is the mountain peak Koćurac (Zladovačka Planina mountain) in the southeastern part of the test area, which is 1558 m a.s.l. (Figures 2 and 3). The study area's vegetation consists of seasonal crops, meadows, pastures, and woodlands. In the northwest and southeast parts of the region are forests. Most forest cover comprises deciduous species (beech, oak, and others), whereas coniferous woods (spruce, fir, and others) comprise a significantly minor portion of the land area. According to geomorphological properties, lowland plains are primarily used for agriculture, where cultivated arable crops such as wheat, maize, and others are planted [22].



**Figure 2.** Vranje and its surroundings—Area of interest. Map created using ESA remote sensing data and © OpenStreetMap contributors open data [22–24].





**Figure 3.** Location of the research area. Created using © OpenStreetMap contributors open data [23].

## 2.2. Satellite Imagery Processing Sentinel-2 Data (Test Data 1)

Materials used in this study primarily contained Sentinel-2 multispectral bands captured on 13 September 2021, processed by ESA (Test Data 1), and obtained using Copernicus Sci-Hub [25]. The Sentinel-2 product was characterised by granules indicative of a specific geographical location. Each granule comprised thirteen unique spectral bands, categorised into three distinct ground resolution levels: 10 m, 20 m, and 60 m. The 10 m bands were: visible Blue (B), Green (G), Red (R), and Near InfraRed (NIR); 20 m bands were Vegetation Red Edge bands, Narrow NIR, and two Short Wave InfraRed (SWIR) bands; and 60 m bands were Coastal Aerosol, Water, Vapour and SWIR Cirrus bands (Table 1) [26].

**Table 1.** List of Sentinel-2 bands used for classification as Test Data 1. This table is created using the data provided in Sentinel-2 MSI User Guide [26] document.

Band	Label	GSD Resolution (m)	Wavelength (nm)
B02	Blue	10	457–522
B03	Green	10	542–577
B04	Red	10	647–682
B05	Red-edge 1	20	697–712
B06	Red-edge 2	20	732–747
B07	Red-edge	20	773–793
B08	Near-infrared (NIR)	10	784–899
B8A	Near-infrared narrow (NIRn)	20	855–875
B10	Shortwave infrared/Cirrus	60	1360–1390
B11	Shortwave infrared 1 (SWIR1)	20	1565–1655

Data preparation included several sub-procedures that were executed on Sentinel-2 bands.

After selecting the AOI, corresponding Sentinel-2 Level-2A data from Copernicus Sci-Hub [25] representing bottom-of-atmosphere reflectance in cartographic geometry was downloaded. Each tile covered  $100 \times 100 \text{ km}^2$  in extent [26]. Furthermore, all 20 m bands (Table 1) were resampled to the spatial resolution of 10 m. The enhancement method that enabled this possibility utilised in scientific research [27–29] employed the Sen2Res plugin [30] in the SNAP [31] software. The Sen2Res tool employed a sophisticated super-resolution technique to merge a band of lower resolution into one of higher resolution while ensuring that the reflectance value remained unaltered. This technique was particularly significant for its ability to delve into the geometric detail information shared among adjacent pixel contents in both the low- and high-resolution bands [30]. The Sen2Res tool, through

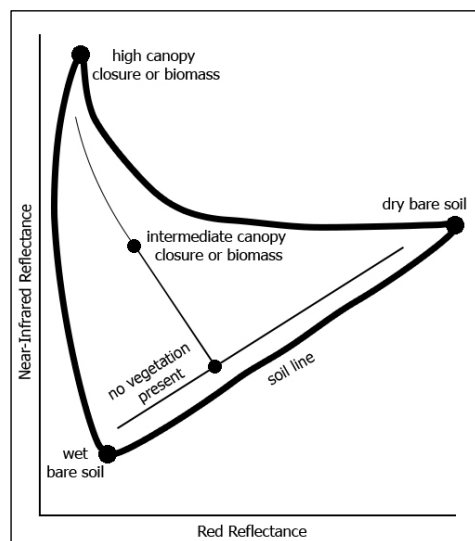
its super-resolution method, achieved its dual objectives by ensuring the uniformity of reflectance values among adjacent pixels in the lower-resolution band and preserving the geometric details of sub-pixel components. This resulted in an enhanced resolution of satellite imagery, with essential details and reflectance values accurately maintained [27–30].

The next step in data preparation included clipping ten Sentinel-2 bands (B02-B08A, B10, and B11) using the AOI polygon in QGIS v.3.28 software [32].

### 2.3. Vegetation Indices (Test Data 2 and Test Data 3)

Certain regions of the electromagnetic spectrum have a particular relationship with healthy green plant canopies. In the visible spectrum, chlorophyll absorbs significant amounts of energy, primarily for photosynthesis. This absorption peaks in the red and blue ranges of the visible spectrum, but chlorophyll reflects the green region, resulting in the typical green colour of most leaves. In addition, the leaf's interior structure greatly reflects the spectrum's near-infrared area [33]. This substantial disparity, especially between the absorbed energy in the red and near-infrared areas of the electromagnetic spectrum, has been the subject of several efforts to construct quantitative measures of vegetation status using remotely sensed images. Vegetation indices (VI's) have been used in various scenarios to evaluate green biomass and as proxies for global environmental change, particularly in drought and land degradation risk assessment [33–35].

VI's can be categorised into three groups: (a) slope based, (b) distance based, and (c) orthogonal transformation vegetation indices [33]. The distribution of vegetation pixels on a two-dimensional graph (or bi-spectral plot) of red versus infrared reflectance should be examined to clarify these differences, where a high portion of biomass is presented with high values in the Infrared band [36] (Figure 4).



**Figure 4.** Red and near-infrared pixel distribution.

- Slope-based vegetation indices (SBVI) are derived by a combination of red and infrared channels (Table 2) and presented as mathematical combinations that emphasise the distinction between vegetation spectral response patterns in the red and near-infrared parts of the electromagnetic spectrum [33];
- Distance-based vegetation indices (DBVI) attempt to neutralise the effect of soil brightness in sparse vegetation areas (Table 2), and they are derived from the Perpendicular Vegetation Index (PVI), which includes the perpendicular distance between each pixel and the soil line [37] (Figure 4). Original PVI is enhanced with three different indices: PVI1 [38], PVI2 [39], and PVI3 [40] to improve its performance;
- Orthogonal transformation vegetation indices (OTVI) are created by applying a transformation on the existing spectral bands to generate a new set of bands that are not in

correlation (Table 2). A green vegetation index band may be generated within this new set of bands [33].

The 23 specific indices (Table 2) are chosen because they capture critical spectral responses of vegetation, such as chlorophyll absorption and near-infrared reflectance, which are vital for distinguishing between forest and non-forest areas. The selection of these indices is thus driven by their effectiveness in enhancing the accuracy of vegetation detection and their ability to capture key spectral characteristics of the vegetation.

**Table 2.** List of vegetation indices used in this research.

Equation No.		Vegetation Index	Equation Adjusted for Sentinel-2 Bands	Group	Author
(1)	AVI-	Ashburn Vegetation Index	$2.0 \times B_{8a} - B_4$	DBVI	[41]
(2)	DVI- <sup>1</sup>	Difference Vegetation Index	$g \times B_{8a} - B_4$	DBVI	[37]
(3)	EVI- <sup>2</sup>	Enhanced Vegetation Index	$G \times \frac{(B_8 - B_4)}{(B_8 + C_1 \times B_4 - C_2 \times B_2 + L) \times (1 + L)}$	SBVI	[42]
(4)	GEMI- <sup>3</sup>	Global Environment Monitoring Index	$\eta \times (1 - 0.25 \times \eta) - \frac{\rho_1 - 0.125}{1 - \rho_1}$ where $\eta = \frac{2 \times (\rho_2^2 - \rho_1^2) + 1.5 \times \rho_2 + 0.5 \times \rho_1}{\rho_2 + \rho_1 + 0.5}$ and $\rho_1 = red_{factor} \times B_4$ $\rho_2 = IRed_{factor} \times B_{8a}$	SBVI	[43]
(5)	GNDVI-	Green Normalized Difference Vegetation Index	$\frac{(B_8 - B_3)}{(B_8 + B_3)}$	SBVI	[44]
(6)	IRECI-	Inverted Red-Edge Chlorophyll Index	$\frac{(B_7 - B_4)}{\left(\frac{B_5}{B_6}\right)}$	SBVI	[45,46]
(7)	MCARI-	Modified Chlorophyll Absorption Ratio Index	$[(B_5 - B_4) - 0.2 \times (B_5 - B_3)] \times \left(\frac{B_5}{B_4}\right)$	SBVI	[47]
(8)	MTCI-	Meris Terrestrial Chlorophyll Index	$\frac{(B_6 - B_5)}{(B_5 - B_4)}$	SBVI	[48]
(9)	NDI45-	Normalised Difference Index	$\frac{(B_5 - B_4)}{(B_5 + B_4)}$	SBVI	[49]
(10)	NDII-	Normalised Difference Infrared Index	$\frac{(B_8 - B_{11})}{(B_8 + B_{11})}$	SBVI	[50]
(11)	NDMI-	Normalised Difference Moisture Index	$\frac{(B_8 - B_{11})}{(B_8 + B_{11})}$	SBVI	[50,51]
(12)	NDVI-	Normalised Difference Vegetation Index	$\frac{(B_8 - B_4)}{(B_8 + B_4)}$	SBVI	[52]
(13)	PSSRA-	Pigment-Specific Simple Ratio	$\frac{B_7}{B_4}$	SBVI	[53]
(14)	PVI- <sup>4</sup>	Perpendicular Vegetation Index	$\left(\frac{1}{\sqrt{a^2 + 1}}\right) \times (B_9 - a \times B_4 - b)$	DBVI	[38,54,55]
(15)	RENDVI-	Red Edge Normalized Difference Vegetation Index	$\frac{(B_6 - B_5)}{(B_6 + B_5)}$	SBVI	[55,56]
(16)	RVI-	Ratio Vegetation Index	$\frac{(B_5)}{(B_4)}$	SBVI	[57]
(17)	S2REP-	Sentinel-2 Red-Edge Position Index	$\frac{705 + 35 \times \left(\frac{B_4 + B_7}{2} - B_5\right)}{B_6 - B_5}$	SBVI	[46,48]
(18)	SAVI- <sup>5</sup>	Soil Adjusted Vegetation Index	$\frac{B_8 - B_4}{B_8 + B_4 + L} \times (1.0 + L)$	SBVI	[58,59]

Table 2. Cont.

Equation No.	Vegetation Index	Equation Adjusted for Sentinel-2 Bands	Group	Author
(19)	TCB- Tasselled Cap— Brightness	$0.3037 \times B_2 + 0.2793 \times B_3 + 0.4743 \times B_4 + 0.5585 \times B_8 + 0.5082 \times B_{10} + 0.1863 \times B_{12}$	OTVI	[60]
(20)	TCG- Tasselled Cap— Green Vegetation Index	$-0.283 \times B_3 - 0.660 \times B_4 + 0.577 \times B_6 + 0.388 \times B_9$	OTVI	[60]
(21)	TCW- Tasselled Cap—wetness	$0.1509 \times B_2 + 0.1973 \times B_3 + 0.3279 \times B_4 + 0.3406 \times B_8 - 0.7112 \times B_{11} - 0.4572 \times B_{12}$	OTVI	[60]
(22)	TVI- Transformed Vegetation Index	$\sqrt{(NDVI) + 0.5}$	SBVI	[61]
(23)	WVG- Water Vapour Grid	$\frac{(B_9)}{(B_{8a})}$	SBVI	[62]

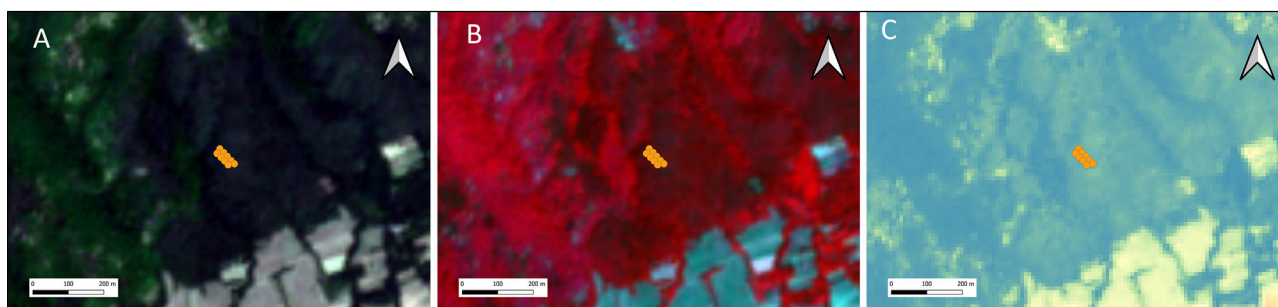
<sup>1</sup>  $g$ —the slope of the soil line. Value used in calculation is 2.4. <sup>2</sup>  $G$ —gain factor = 2.5,  $C1$ —constant = 6,  $C2$ —constant = 7.5, and  $L$ —Soil adjustment factor = 1. <sup>3</sup> Values used in calculation are Redfactor = 1, IRRedfactor = 1. <sup>4</sup>  $a$  is slope of the soil line (the default value is 0.3), and  $b$  is gradient of the soil line (the default value is 0.5) [55]. <sup>5</sup>  $L$  changes with the reflectance characteristics of the soil. In situations with very low vegetation, a  $L$  factor of 1.0 is recommended, 0.5 for moderate densities, and 0.25 for high densities [33].

#### 2.4. Samples Collection

A sample set including 5433 individual points was defined to fulfil research objectives. All points were classified as “forest” (class 1) or “not forest” (class 0). The “forest” class included all forest varieties, whereas the “not forest” class included all non-forest geospatial elements (meadows, fields, rivers, lakes, communications, facilities, and others).

A total of 5433 points were defined in the sampling process, of which 2509 points were classified in class 1—“forest”, and 2924 points were classified in class 0—“not forest” (Figure 5). Detailed VHR (very high resolution) aerial photos and satellite images were used to define the samples as precisely as possible.

A comprehensive analysis was conducted on the samples to guarantee the exactness of the classifications. A substantial fraction, as much as 10%, underwent rigorous on-site re-evaluation, enhancing the validation procedure and confirming the precision of the results. In certain instances, when the sample class could not be determined from the accessible satellite and aerial imagery, instant field-based identification and verification were performed to obtain the most precise dataset.



**Figure 5.** Sampling position of forest class (orange dots): (A) RGB colour composite, (B) NIRR False colour composite, (C) NDVI [26]. Copyright: Authors, contains modified Copernicus Sentinel data 2021.

#### 2.5. Training and Test Data Definition

Applying the machine learning process required the prior definition of a Training and Test Data set. Accordingly, these sets were defined for 5433 sampled points and corre-

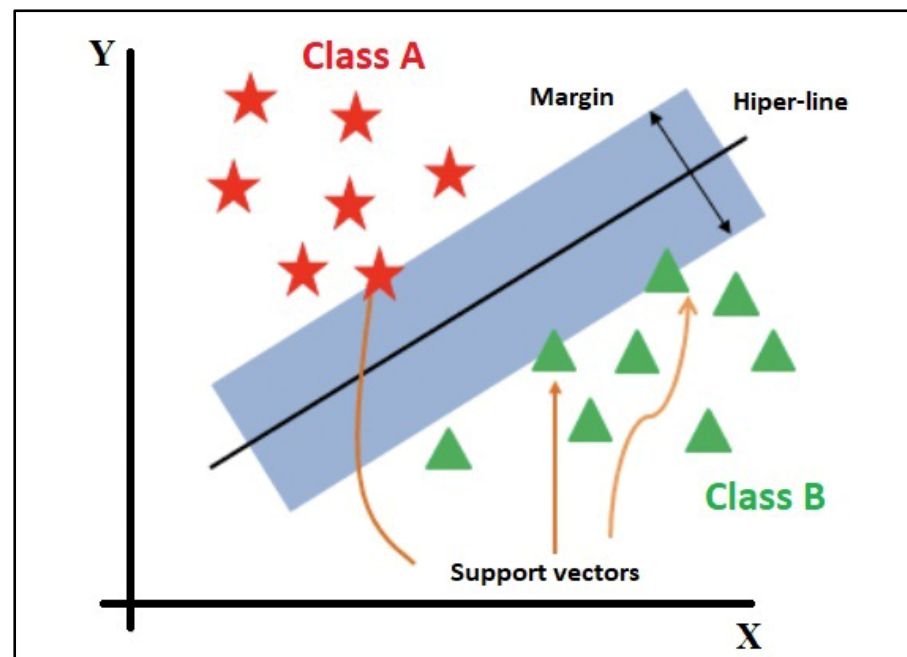


sponding values of ten Sentinel-2 10 m bands (marked bands in Table 1). The 3622 points (70%) were determined by random selection, and the remaining 1811 points (30%) were engaged for accuracy assessment out of 5433 points in total, for which classes were defined to use in the learning (training) process. The 70:30 split for training and testing data in machine learning models is a widely accepted standard. This ratio balances optimising the model's learning capacity and ensuring sufficient data for validation. It reduces the risk of overfitting, where the model performs poorly on new data, and underfitting, where the model underperforms due to insufficient learning. The 30% test set typically offers statistically significant performance measures [63–65]. However, the optimal split may vary depending on the specific dataset and problem, and techniques like cross-validation can be employed to utilise the data effectively.

The entire data set was divided into two primary groups considering the research goal and providing different “origin” parameters. The first group of the test data (Test Data 1) consisted of values based on ten Sentinel-2 bands, and the second group (Test Data 2) consisted of data obtained from 23 vegetation indices (Figure 1). The “Third” data group was derived using Test Data 2 values that positively impacted the dataset's accuracy.

### 2.6. Support Vector Machines (SVM) Algorithm

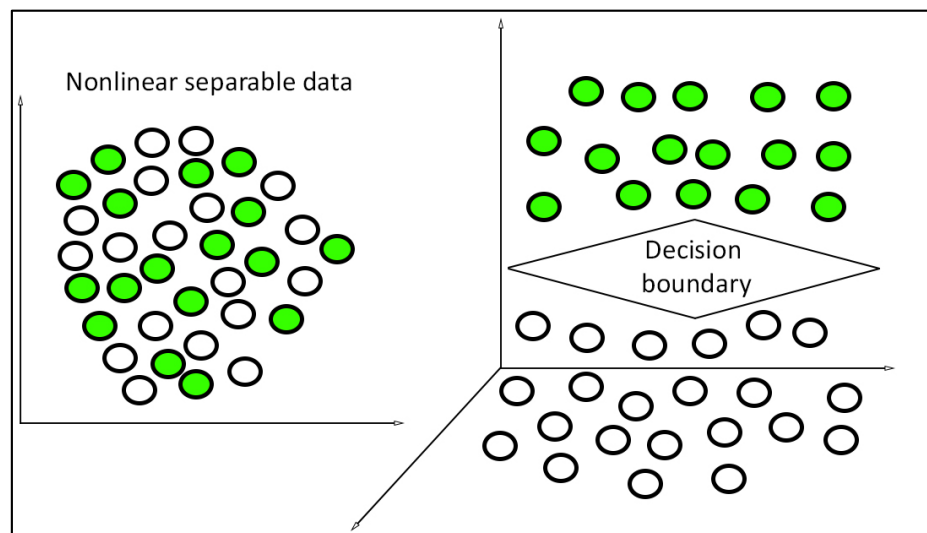
Support Vector Machines (SVM) is a popular machine learning algorithm for classification and regression problems [66]. In this study, SVM is employed for forest detection and classification. SVM stands for support vector machine and is a technique for supervised machine learning capable of performing classification, regression, and even outlier identification. SVM is a linear binary classifier in its most basic form, and a support vector classifier (SVC) is defined for that purpose. This kind of classifier finds a single border between two classes. The linear SVM assumes that the multidimensional data are linearly separable in the input space (Figure 6) [66,67].



**Figure 6.** Linear SVM working method.

Different classes of data samples could not always be linearly separated from one another and often overlap (Figure 7). As a result, the linear SVM could not guarantee high accuracy while categorising this data and required certain adjustments [68,69]. To circumvent the constraints imposed by linear SVM, Cortes and Vapnik developed two new methods: the soft margin and the kernel trick [70]. To handle nonlinearly separable data (Figure 7), the soft margin method to SVM optimisation might have extra variables—also

known as slack variables—added to the optimisation process. The kernel trick projected non-linear data onto a higher dimension space to facilitate the classification of the data in situations where it may be linearly separated by a plane [70]. The SVM module from Python scikit-learn package [71,72] was employed for the data classification in this research.



**Figure 7.** Non-linear SVM kernel trick.

#### 2.6.1. Radial Basis Function (RBF)

The RBF kernel of the SVM learning algorithm was used for the categorisation of Test Data within the scope of forest cover detection, which is presented in Equation (24) [71]:

$$\exp(-\gamma\|x - x'\|^2) \quad (24)$$

where  $\gamma$  is specified by parameter gamma and must be a positive value (greater than 0), and  $\|x - x'\|^2$  is the square of the Euclidean distance between the points  $x$  and  $x'$ .

The performance of SVMs, particularly those employing the RBF kernel, hinged significantly on the choice of parameters  $C$  and gamma ( $\gamma$ ), the fundamental elements defining the SVM algorithm. This essential selection process, known as hyperparameter tuning, optimised the model's performance.

Parameter  $C$ , the cost or regularisation parameter, mediated the delicate balance between reducing training data error and mitigating model complexity to evade overfitting. A smaller  $C$  fostered a broader margin, tolerating more misclassifications yet potentially yielding a more straightforward, less overfitted decision function. On the contrary, a larger  $C$  pursued a tighter margin and fewer misclassifications, which might engender a more complex model susceptible to overfitting [72]. The gamma parameter, integral to the RBF kernel, designated the reach of a single training example's influence. An example's influence was widespread if low, resulting in a broad, smooth decision boundary. Conversely, a high gamma suggested a limited influence, creating an irregular decision boundary that may capture finer detail and potential noise in the dataset [70]. This parameter could also be seen as the inverse of the standard deviation of the Gaussian function used in the RBF kernel, emphasising the closeness of data points as a similarity measure.

The quest for optimal  $C$  and gamma values typically necessitates iterative testing of various parameter combinations, employing grid and randomised search techniques. The preferred model produced the highest mean test accuracy across all iterations. In a  $k$ -fold dataset division, each unique pair of  $C$  and gamma parameters trains the model  $k$  times, utilising different folds as the test set in each instance [71]. Bishop's influential text *Pattern Recognition and Machine Learning* underscored this procedure's import: "The RBF kernel has two parameters:  $C$  and gamma. The optimum settings for these parameters are data-

dependent and must be determined through experimentation. Typically, a range of values are evaluated on a validation set, and the best performance parameters are chosen" [73]. While grid and randomised search methods are standard and straightforward, it is crucial to correctly carry out the validation process to prevent model overfitting to training data, potentially compromising the model's generalisation ability. More sophisticated methods, such as Bayesian optimisation, might offer improved results in some instances [74,75]. The optimisation of the SVM parameters process involves tuning the parameters to find the optimal values that yield the best performance. The grid search method used in this study is a commonly used technique for parameter optimisation. A grid search algorithm is a method for optimising the parameters of a machine learning model by exhaustively searching through a manually specified subset of the hyperparameter space. The algorithm evaluates the model's performance on a validation set and adjusts the hyperparameters to maximise the performance metric. The algorithm then repeats this process until it finds the optimal hyperparameter combination [76]. For the SVM, the grid search method involves selecting a range of values for the parameters C and gamma. The range of these parameters is usually chosen based on the problem at hand and the nature of the data. The grid search method then trains an SVM for each pair of (C, gamma) values in the Cartesian product of the two ranges. The optimal choice is the pair with the best cross-validation accuracy [77,78].

For the cost parameter C, we explored values in the range of 0.1–1000 with a step size of 10. For the gamma parameter, we explored values in the range of 0.1–10 with a step size of 0.1 (Table 3). These ranges were chosen based on common practice [72,79] and the specific characteristics of our dataset. The performance of the SVM model with each combination of parameters was evaluated using cross-validation. Specifically, we used a 10-fold cross-validation process, which involved splitting the dataset into ten subsets and then training and testing the model 10 times, each time with a different subset as the test set.

**Table 3.** Part of Full Grid Search Results for SVM Hyperparameters C and Gamma.

Iteration	C	Gamma	Accuracy
1	0.1	0.1	0.6
2	10.1	0.2	0.62
3	20.1	0.3	0.64
...	...	...	...
50	490.1	2.9	0.93
51	500.1	3	0.95
...	...	...	...
100	990.1	9.9	0.9
101	1000.1	10	0.88

The combination of parameters that resulted in the highest cross-validation score was selected as the optimal parameters for our SVM model. In our case, the optimal values were C = 500 and gamma = 3. These values indicated a relatively high penalty for misclassification (C = 500), and a low influence range of the samples (gamma = 3), implying that the model was complex and may have high variance.

The RBF kernel is a popular choice for SVM due to its locality and finite response across the entire range of the real  $x$ -axis. It is a good choice when there is no prior knowledge about the data. The RBF kernel performs a non-linear mapping of samples to a higher-dimensional space, effectively handling situations where the relationship between class labels and attributes is not linear, unlike the linear kernel [80].

### 2.6.2. Utilising SVM with SVC in Python Programming

In this research, the Python programming language [81] executed the primary process, which involved loading prepared vector and raster data through a direct connection or connection to an SQL database. Python (3.6.10) with the following packages was used:

scikit-learn 0.24.2 [71,82], SVM 0.1.0 [83], GDAL 3.0.4 [84], rasterio 1.1.4 [85], pyodbc 4.0.32 [86], and NumPy [87]. In addition to the programming language, ArcGIS Pro [88] and SQL Express [89] displayed the machine learning results and archived sampling data. The Python code utilised for the classification of satellite imagery in this study is presented in Listing A1 of Appendix A.

### 2.7. Accuracy Assessment

The accuracy assessment procedure for a Support Vector Machine (SVM) model, as implemented in the Scikit-learn library in Python, involved several key steps.

Firstly, the dataset was divided into training and testing sets. We used 70% of the data allocated for training the model, and the remaining 30% was reserved for testing the model's performance. This division was vital to ensure that the model was not tested on the same data it was trained on, which could have led to overfitting and a misleadingly high accuracy score [90,91].

The next step was to train the SVM model using the training data. In Scikit-learn, this was accomplished by creating an instance of the SVM classifier and fitting it to the training data. The SVM algorithm attempted to find a hyperplane in an N-dimensional space that distinctly classified the data points.

Once the model was trained, it could predict the unseen Test Data. The model used the hyperplane determined during training to classify the new data points.

Finally, the model's accuracy was assessed by comparing the predicted values to the actual values from the Test Data. The accuracy score was a standard metric for this purpose, which calculated the proportion of correct predictions out of the total predictions. However, it is essential to note that relying only on accuracy as a measure may not offer a comprehensive evaluation of the model's effectiveness [92–95]. Other metrics, such as precision, recall, F1 score, or area under the ROC curve, might also be considered, depending on the specific use case [70,71,96]. Furthermore, it is vital to underscore the significance of evaluating other quality dimensions in conjunction with thematic accuracy. This principle could also be relevant to the assessment of SVM models [91].

## 3. Results

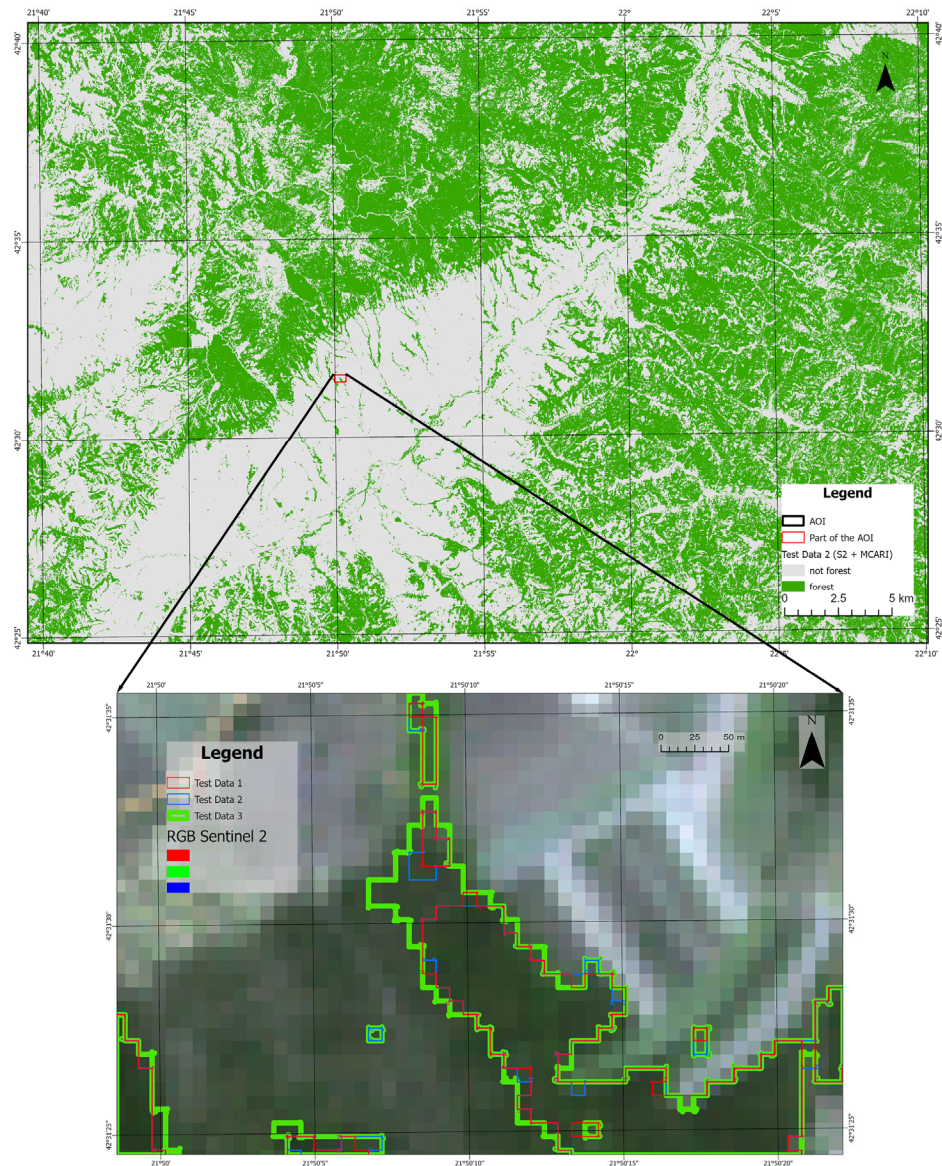
Based on the results presented in Table 4 and Figure 8, it can be observed that the detected forest area increased as the dataset changes. Test Data 1 resulted in a detected area of 700.81 km<sup>2</sup> (57.54%), whereas Test Data 2 yielded a slightly larger area of 705.06 km<sup>2</sup> (57.89%). However, the largest detected forest area of 706.30 km<sup>2</sup> (57.99%) was recorded with Test Data 3. These findings indicated that Test Data 3 provided additional features that enhanced the accuracy of forest detection, making it the most effective dataset for this purpose. The consistent increase in detected forest areas across the different data variations further validated the reliability of Test Data 3 in accurately identifying forested regions.

**Table 4.** Impact of Data Variations on Detected Forest Areas.

Data Set	Detected Forest Area (km <sup>2</sup> )	Detected Forest Area (%)
Test Data 1	700.81	57.54
Test Data 2	705.06	57.89
Test Data 3	706.30	57.99

The accuracy evaluation of the acquired results included assessing the classification accuracy performed on the Test Data 1 classification (ten Sentinel-2 bands) in the first step. In the second step, an accuracy assessment was performed for the classification, where individual vegetation indices (Test Data 2) were added to Sentinel-2 bands (Table 5). In the final step, an accuracy assessment was performed for the classification, for which index groups (Test Data 3) were added to Test Data 1 as input data for the classification process (Table 6).





**Figure 8.** AOI Test Data 2 result and Detected Forest Areas across Different Datasets: Visualising part of the Area of Interest. Copyright: Authors, contains modified Copernicus Sentinel data 2021.

The overall accuracy of classification performed using Test Data 1 was 98.18%. This result was a reference point for comparison with other classification results.

**Table 5.** The overall classification accuracy is performed using Test Data 1 and each index from Test Data 2 individually. S. No. column presents the accuracy assessment quality group.

Index Reference to Table 2 Equation No.	S. No.	Data Combinations (Test Data 1 + Test Data 2)	Accuracy (%) (Test Data 1 + Test Data 2)
(7)	1	Test Data 1 + MCARI	98.56
(15)		Test Data 1 + RENDVI	98.56
(9)	2	Test Data 1 + NDI45	98.40
(5)	3	Test Data 1 + GNDVI	98.29
(10)	4	Test Data 1 + NDII	98.23

Table 5. Cont.

Index Reference to Table 2 Equation No.	S. No.	Data Combinations (Test Data 1 + Test Data 2)	Accuracy (%) (Test Data 1 + Test Data 2)
(1)	5	Test Data 1 + AVI	98.18
(6)		Test Data 1 + IRECI	98.18
(11)		Test Data 1 + NDMI	98.18
(12)		Test Data 1 + NDVI	98.18
(18)		Test Data 1 + SAVI	98.18
(19)		Test Data 1 + TCB	98.18
(22)		Test Data 1 + TVI	98.18
(20)	6	Test Data 1 + TCG	98.12
(21)		Test Data 1 + TCW	98.12
(23)		Test Data 1 + WVG	98.12
(2)	7	Test Data 1 + DVI	98.07
(3)		Test Data 1 + EVI	98.07
(8)		Test Data 1 + MTCI	98.07
(13)		Test Data 1 + PSSRA	98.07
(14)		Test Data 1 + PVI	98.07
(17)	8	Test Data 1 + S2REP	98.01
(4)	9	Test Data 1 + GEMI	97.96
(16)	10	Test Data 1 + RVI	97.90

Table 6. The overall classification accuracy is performed using Test Data 1 and Test Data 3 (VI's groups). "Test Data 1" and "Test Data 1 + mcari" are set as reference points.

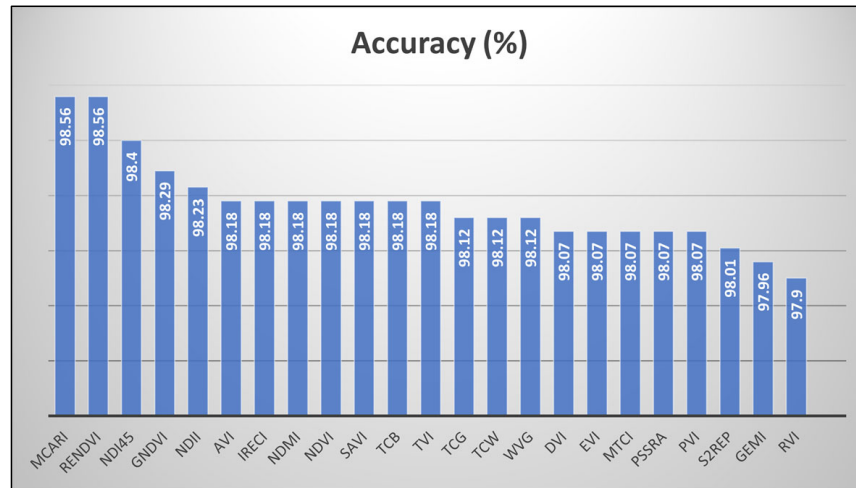
Data Combinations	Accuracy (%)
Test Data 1	98.18
Test Data 1 + mcari	98.56
Test Data 1 + mcari, rendvi	98.67
Test Data 1 + mcari, rendvi, ndi45	98.73
Test Data 1 + mcari, rendvi, ndi45, gndvi	98.79
Test Data 1 + mcari, rendvi, ndi45, gndvi, ndii	99.01

The accuracy varied when the SVM classification procedure included individual vegetation indices from Test Data 2. The improved accuracy ranged from 98.56 to 98.23% (Table 5). Five indices increased the classification accuracy (S. No. 1–4), whereas the other indices had no positive effect (S. No. 5) or had a negative effect (S. No. 6–10) on classification accuracy (Table 5 and Figure 9).

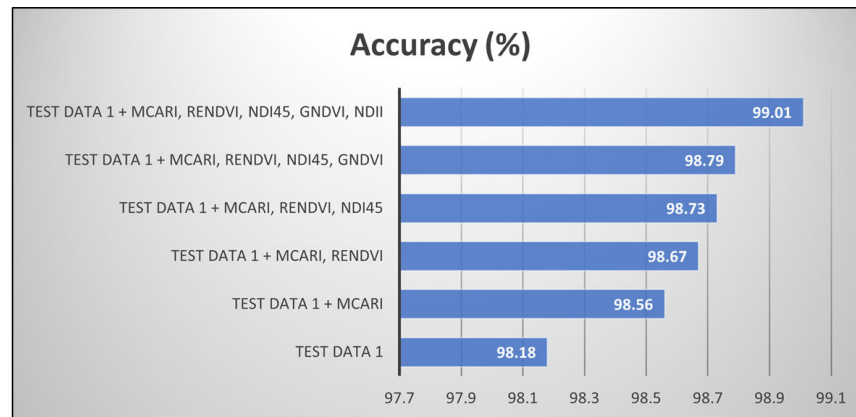
The classification results of Test Data 1 and 2 combinations gave a new potential to the multispectral analysis of satellite imagery. These positive results confirmed the initial hypothesis of this research, so the next step consisted of creating groups of vegetation indices (Test Data 3).

The first Test Data 3 group was created using five vegetation indices that enhanced the accuracy of Test Data 1 (Table 5 Serial Numbers 1–4).

Test Data 3 combinations were added to Test Data 1 and included in the SVM procedure. The accuracy assessment of these data combinations indicated that the improvement in classification accuracy with the vegetation indices was not unnoticeable or negligible. The highest accuracy of 99.01 % was achieved by combining Test Data 1 with five vegetation indices: MCARI, RENDVI, NDI45, GNDVI, AND NDII (Table 6 and Figure 10).

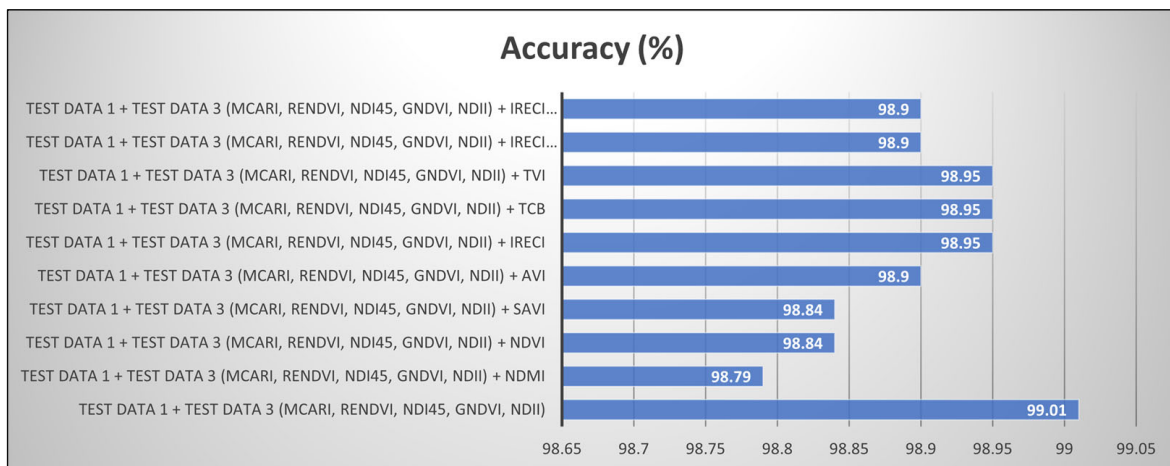


**Figure 9.** Graphical representation of the comprehensive accuracy achieved through classification, utilising Test Data 1 in conjunction with each index from Test Data 2.



**Figure 10.** The overall accuracy chart of classification performed using Test Data 1 and Test Data 3.

Furthermore, research was carried out on a data set that included indices with S. no. 5 from Table 5. Six, seven, and eight (one, two, or three indices added to Test Data 3) different indices were grouped (Table 7). The increasing accuracy trend caused by VI’s influence grew (Table 6) until further addition of the index had no significant effect (Figure 11).

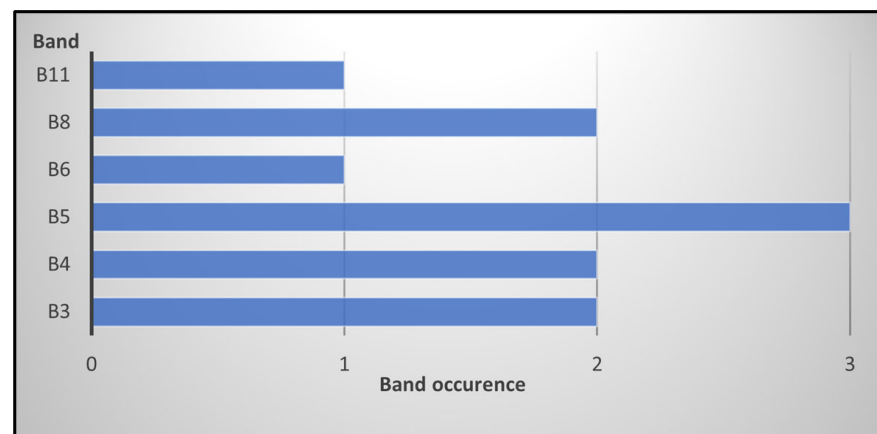


**Figure 11.** The overall accuracy chart of classification is performed using Test Data 1 + Test Data 3 + additional VI’s marked with S. No. 5 in Table 5.

**Table 7.** The overall classification accuracy is performed for: Test Data 1 + Test Data 3 + additional VI's marked with S. No. 5 in Table 5.

Data Combinations	Accuracy (%)
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii)	99.01
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + ndmi	98.79
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + ndvi	98.84
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + savi	98.84
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + avi	98.90
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + ireci	98.95
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + tcb	98.95
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + tvi	98.95
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + ireci + tcb	98.90
Test Data 1 + Test Data 3 (mcari, rendvi, ndi45, gndvi, ndii) + ireci + tcb + tvi	98.90

Conclusively, the most influential bands used in Test Data 3 (MCARI, RENDVI, NDI45, GNDVI, NDII) that improved classification accuracy are presented in Figure 12. The most influential bands that positively affected the accuracy of the classification were Band 5 (Red-edge 1), Band 3 (Green), Band 4 (Red), and Band 8 (NIR). Band 5 (Red-edge 1) appeared three times, whereas Bands 3 (Green), 4 (Red), and 8 (NIR) each appeared twice.

**Figure 12.** The most significant bands that improved the classification accuracy.

#### 4. Discussion

This research presents a novel approach to forest detection and classification using remote sensing data and machine learning techniques, specifically the Support Vector Machine (SVM) algorithm. The study focuses on integrating vegetation indices and multispectral bands as input parameters for the SVM classification. Including these indices, which capture specific relationships between spectral bands and vegetation properties, significantly enhanced forest detection accuracy.

The initial classification, which used Sentinel-2 bands as input parameters, yielded an accuracy of 98.18%. However, when individual vegetation indices were incorporated, the accuracy ranged from 98.23% to 98.56%. The accuracy was further improved by selecting the best-performing vegetation indices to 99.01%. The improvement in accuracy demonstrated the effectiveness of vegetation indices in enhancing the quality of forest detection. While the enhancement in accuracy may have appeared marginal at less than 1%, it was crucial to interpret this advancement within its broader context and implications. In remote sensing and forest detection, even minor increments in accuracy could yield substantial outcomes [10,97,98]. For example, a 1% augmentation in accuracy could equate to precisely



identifying several square kilometres of forest that might otherwise be inaccurately classified. This precision can be pivotal for deforestation monitoring, conservation planning, or carbon stock estimation applications. Consequently, the endeavour invested in boosting the model's accuracy, despite yielding a seemingly modest improvement, could be warranted considering the potential ramifications of these significant applications.

The study also emphasised the importance of carefully selecting and evaluating vegetation indices for specific classification tasks. Some indices positively impacted the classification accuracy, whereas others had no effect or even degraded the accuracy. The five indices that contributed the most significantly to the increase in classification accuracy were MCARI, RENDVI, NDI45, GNDVI, and NDII.

The SVM, a supervised machine learning technique, identifies an optimal hyperplane that separates different classes based on the training data. Selecting appropriate parameters, such as the gamma and C coefficients, proves crucial for achieving accurate results. The C coefficient is set at 500 through empirical determination, whereas the gamma coefficient is established at 3. The findings of this study align with other research in the field, such as [1,3–5,18–21], which also explored the use of remote sensing data and machine learning for vegetation classification and monitoring. These studies underscore the potential of remote sensing techniques for assessing deforestation, land cover changes, and monitoring vegetation degradation.

This study distinguished itself through its unique emphasis on the SVM algorithm and binary classification for forest detection.

In contrast, other studies, like those by Li et al. (2019) [99] and Baldeck et al. (2015) [100], employed different methodologies. Li et al.'s study utilised a two-stage convolutional neural network (TS-CNN) for oil palm tree detection in a large-scale study area in Malaysia. Their approach achieved a high average F1-score of 94.99% in their study area. However, their method required very high-resolution images. It did not consider the features of the plantation region, which may have limited its applicability in different contexts or regions with lower-resolution images.

On the other hand, using airborne imaging spectroscopy data, Baldeck et al.'s study focused on identifying individuals belonging to three specific canopy tree species amidst a varied assortment of tree and liana species on Barro Colorado Island, Panama. The researchers employed binary SVM and biased SVM techniques to evaluate their effectiveness in distinguishing pixels associated with a particular focal species. Their methodology demonstrated excellent precision in identifying the focal species, with pixel-level producer accuracies ranging from 94% to 97% for the three species in focus.

Furthermore, field validation of the predicted crown objects confirmed user accuracies ranging from 94% to 100%. However, their study was limited to three focal species and required high-resolution imaging spectroscopy data. Furthermore, the study by Nasiri et al. [101] offered a different perspective on using machine learning in environmental studies. Their research focused on mapping forest canopy cover (FCC) in Mediterranean oak forests using Sentinel-1 and Sentinel-2 time series. They employed Support Vector Machine (SVM), Random Forest (RF), and Classification and Regression Tree (CART) machine learning models. Their results showed that SVM outperformed RF and CART in terms of accuracy, irrespective of data density and integration. However, their study was focused on FCC mapping, which, while related, was a different task from binary classification for forest detection. Furthermore, their study required the integration of Sentinel-1 and Sentinel-2 spectral-temporal metrics, which may have limited its applicability in regions with different satellite coverage or data availability.

Alternatively, another study focused on land use/land cover (LULC) mapping using satellite time series [102]. The authors used the Random Forest classifier to produce accurate LULC maps, similar to the approach used in this paper. However, their study focused on LULC mapping, a task different from binary classification for forest detection. Moreover, their study required spectral-temporal metrics from satellite time series, which may have limited its applicability in regions with different satellite coverage or data avail-

ability. Furthermore, a comparison of earlier studies incorporating vegetation indices as supplementary bands to multispectral satellite bands in machine learning applications will be conducted. Vegetation indices' effectiveness in distinguishing between different vegetation types, as demonstrated in Fletcher's (2016) research [18], was echoed in the present study's application to forest detection. This research, however, extended the application of these indices to a larger scale, highlighting their potential in real-world scenarios. The automatic segmentation method for vegetation detection in precision agriculture proposed by Turhal (2022) [19] shared methodological similarities with the present study, particularly in using vegetation indices. The focus, however, diverged, as this research was centred on forest detection. The employed machine learning algorithm also differed with the present study utilising support vector machines (SVM), which was proven effective in the given context. Contrasting with the Li et al.'s (2021) paper, which used the extreme gradient boosting method to predict vegetation growth [20], this research employed SVM for forest detection. Despite the different methods and applications, both studies underscored the versatility of machine learning and vegetation indices in environmental studies. Sener and Arslanoglu (2019) emphasised the selection of suitable Sentinel-2 bands and vegetation indices for crop classification [21], which aligned with the present study's use of Sentinel-2 multispectral bands and various vegetation indices. However, this research expanded on this by demonstrating the effectiveness of these tools in a different context, namely, forest detection.

While all these studies demonstrated the power and versatility of machine learning algorithms in environmental studies, they differed from this paper in their focuses and methodologies. This paper's unique emphasis on binary classification for SVM forest detection set it apart. It augmented the progressively increasing academic research on the application of machine learning algorithms in forest management. Furthermore, this study's use of the SVM algorithm for forest detection was not limited to a few focal species and did not require high-resolution imaging spectroscopy data. The flexibility of the SVM algorithm allowed it to handle a broader range of forest types and conditions, making it a more versatile tool for forest detection.

Furthermore, this study's second focus on defining and optimising the SVM parameters added another layer of precision and adaptability to the model. This approach contributed to the field and facilitated new possibilities for future research. Compared to the studies above, the present research offered several distinct advantages. It demonstrated the effectiveness of machine learning and remote sensing in a large-scale, real-world application, providing valuable insights for environmental resource management. The achieved high classification accuracy underscored the potential of the used approach. Furthermore, the methodology and findings could be applied to other regions and ecosystems, contributing to the preservation and conservation of forest ecosystems globally.

Although the study's findings are promising, it is essential to note that they are specific to the study area and datasets used. Generalising these findings to other regions and datasets should be performed cautiously, considering the variability of vegetation types and environmental conditions. Future research could explore the application of the proposed methodology to larger study areas and different types of ecosystems. Additionally, integrating other data sources, such as LiDAR or hyperspectral imagery, could further enhance the accuracy and detail of vegetation classification. Evaluating the performance of other machine learning algorithms and comparing them with SVM could also be beneficial for identifying the most suitable approach for specific classification tasks.

## 5. Conclusions

The utilisation of machine learning for vegetation identification through satellite imagery is a notable instance of its application. Typically, multispectral images captured by satellite sensors are employed as input parameters. This study is initiated with a hypothesis: suitable vegetation indices, used with multispectral bands, could serve as effective input parameters for machine learning classification.

The study was performed with three sets of input parameters—the initial set comprised ten bands from the Sentinel-2 mission, the second set included 23 distinct vegetation indices, and the final set was established after analysing the 23 indices to select the most effective vegetation indices for inclusion in the input parameters. Following the analysis, five vegetation indices were chosen. The first set of input parameters achieved an accuracy of 98.18%, a creditable outcome. The accuracy increased to 98.56% with the second set of input parameters and further to 99.01% with the third set. The increased accuracy achieved by including the vegetation index in the input parameters confirmed that this approach could improve the quality of machine learning classification results.

Conducted in the southeastern region of Serbia, this study provided valuable insights into applying machine learning for forest detection within a specific geographical context. The study area's diverse vegetation and terrain characteristics posed a significant challenge for accurate forest classification. However, the accuracy of the classification process could be improved by incorporating suitable vegetation indices.

MCARI, RENDVI, NDI45, GNDVI, and NDII were particularly effective in enhancing classification accuracy among the chosen vegetation indices. These indices captured critical spectral responses of vegetation, such as chlorophyll absorption and near-infrared reflectance, which are vital for distinguishing between forest and non-forest areas.

This research underscored the importance of incorporating vegetation indices into machine learning classification for forest detection. The results showed that the accuracy of forest classification significantly improved when these indices were combined with multispectral bands. These findings had significant implications for environmental resource management, emphasising the potential of integrating advanced technologies like machine learning and remote sensing to enhance forest ecosystem preservation and conservation.

Future research should explore the applicability of these findings in other geographical regions and further investigate other vegetation indices to improve the accuracy of forest detection and monitoring processes.

**Author Contributions:** Conceptualisation, Z.S.; methodology, I.P.; resources, S.B.; software, B.V.; supervision, D.Đ.; validation, J.M.J.; visualisation, R.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This paper is written as part of Project 1.23/2023 of the Ministry of Defense and the Serbian Army. © OpenStreetMap contributors' data are available under the Open Database Licence.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Listing A1:** Utilised Python code for SVM classification.

```
from sklearn import svm
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import rasterio
import numpy as np

# raster training zones
reclassified_raster_path = '... training.tif'
```

```
# Sentinel 2 MS bands and indices
channel_paths = [
'... _B2.tif',
'... _B3.tif',
'... _B4.tif',
'... _B8.tif'
# '... xxxx.tif' other bands and indices
]

# Loading Sentinel 2 MS bands
channel_data = []
for channel_path in channel_paths:
with rasterio.open(channel_path) as src:
channel_data.append(src.read(1))

# Loading the target variable
with rasterio.open(reclassified_raster_path) as src:
y = src.read(1)

# Reshaping the data into a format suitable for SVM
X = np.stack(channel_data, axis=-1)
X = X.reshape(-1, X.shape[-1])
y = y.ravel()

# Ignoring NODATA pixels
nodata_mask = y != -9999
X = X[nodata_mask]
y = y[nodata_mask]

# Splitting the data into training and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.30, random_state=42)

# Creating an SVM classifier
classifier = svm.SVC(kernel='rbf', C=500, gamma=3)

# Training the classifier on the training set
classifier.fit(X_train, y_train)

# Predicting on the test set
y_pred = classifier.predict(X_test)

# Calculating accuracy
accuracy = accuracy_score(y_test, y_pred)

# Saving the accuracy into a text file
with open('... classified_accuracy.txt', 'w') as f:
f.write('Accuracy of the model: ' + str(accuracy))

# Classifying the entire image
X_full = np.stack(channel_data, axis=-1)
classified_data = classifier.predict(X_full.reshape(-1, X_full.shape[-1]))

# Returning classified_data to its original shape
classified_data = classified_data.reshape(X_full.shape[:-1])
```



```
# Saving the classified image
classified_raster_path = '... classified.tif'
with rasterio.open(channel_paths[0]) as src:
    profile = src.profile
    profile.update(count=1, dtype=rasterio.uint8, compress='lzw', nodata=0)

with rasterio.open(classified_raster_path, 'w', **profile) as dst:
    dst.write(classified_data.astype(rasterio.uint8), 1)
```

## References

- Zhu, M.; Zhang, J.; Zhu, L. Variations in Growing Season NDVI and Its Sensitivity to Climate Change Responses to Green Development in Mountainous Areas. *Front. Environ. Sci.* **2021**, *9*, 678450. [CrossRef]
- Blackman, A. Evaluating Forest Conservation Policies in Developing Countries Using Remote Sensing Data: An Introduction and Practical Guide. *For. Policy Econ.* **2013**, *34*, 1–16. [CrossRef]
- Potić, I.; Mihajlović, L.M.; Šimunić, V.; Čurčić, N.B.; Milinčić, M. Deforestation as a Cause of Increased Surface Runoff in the Catchment: Remote Sensing and SWAT Approach—A Case Study of Southern Serbia. *Front. Environ. Sci.* **2022**, *10*, 682. [CrossRef]
- Potic, I.; Curcic, N.; Radovanovic, M.; Stanojevic, G.; Malinovic-Milicevic, S.; Yamashkin, S.; Yamashkin, A. Estimation of Soil Erosion Dynamics Using Remote Sensing and Swat in Kopaonik National Park, Serbia. *J. Geogr. Inst. Jovan Cvijic SASA* **2021**, *71*, 231–247. [CrossRef]
- Woodcock, C.E.; Allen, R.; Anderson, M.; Belward, A.; Bindschadler, R.; Cohen, W.; Gao, F.; Goward, S.N.; Helder, D.; Helmer, E.; et al. Free Access to Landsat Imagery. *Science* **2008**, *320*, 1011. [CrossRef] [PubMed]
- Lausch, A.; Erasmi, S.; King, D.; Magdon, P.; Heurich, M. Understanding Forest Health with Remote Sensing -Part I—A Review of Spectral Traits, Processes and Remote-Sensing Characteristics. *Remote Sens.* **2016**, *8*, 1029. [CrossRef]
- Montzka, C.; Bayat, B.; Tewes, A.; Megen, D.; Vereecken, H. Sentinel-2 Analysis of Spruce Crown Transparency Levels and Their Environmental Drivers After Summer Drought in the Northern Eifel (Germany). *Front. For. Glob. Chang.* **2021**, *4*, 667151. [CrossRef]
- Kaplan, G.; Avdan, U. Algorithm for snow monitoring using remote sensing data. *ANADOLU Univ. J. Sci. Technol. A-Appl. Sci. Eng.* **2017**, *18*, 238. [CrossRef]
- Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing. *Sensors* **2020**, *20*, 6442. [CrossRef] [PubMed]
- Housman, I.W.; Chastain, R.A.; Finco, M.V. An Evaluation of Forest Health Insect and Disease Survey Data and Satellite-Based Remote Sensing Forest Change Detection Methods: Case Studies in the United States. *Remote Sens.* **2018**, *10*, 1184. [CrossRef]
- Chen, W.; Hu, X.; Chen, W.; Hong, Y.; Yang, M. Airborne LiDAR Remote Sensing for Individual Tree Forest Inventory Using Trunk Detection-Aided Mean Shift Clustering Techniques. *Remote Sens.* **2018**, *10*, 1078. [CrossRef]
- Kotsiantis, S.B.; Zaharakis, I.D.; Pintelas, P.E. Machine Learning: A Review of Classification and Combining Techniques. *Artif. Intell. Rev.* **2006**, *26*, 159–190. [CrossRef]
- Smith, S. Metrics for Decision Making. *Pract. Tour. Res.* **2017**, *2017*, 154–184. [CrossRef]
- Joshi, A.V. Essential Concepts in Artificial Intelligence and Machine Learning. In *Machine Learning and Artificial Intelligence*; Springer: Cham, Switzerland, 2023; pp. 7–20. [CrossRef]
- Samuel, A.L. Some Studies in Machine Learning Using the Game of Checkers. *IBM J. Res. Dev.* **1959**, *3*, 210–229. [CrossRef]
- Awad, M.; Khanna, R. Machine Learning. In *Efficient Learning Machines*; Awad, M., Khanna, R., Eds.; Apress: Berkeley, CA, USA, 2015; pp. 1–18.
- Drobnjak, S.; Stojanović, M.; Djordjević, D.; Bakrač, S.; Jovanović, J.; Djordjević, A. Testing a New Ensemble Vegetation Classification Method Based on Deep Learning and Machine Learning Methods Using Aerial Photogrammetric Images. *Front. Environ. Sci.* **2022**, *10*, 702. [CrossRef]
- Fletcher, R.S. Using Vegetation Indices as Input into Random Forest for Soybean and Weed Classification. *Am. J. Plant Sci.* **2016**, *7*, 2186–2198. [CrossRef]
- Turhal, U.C. Vegetation Detection Using Vegetation Indices Algorithm Supported by Statistical Machine Learning. *Environ. Monit. Assess.* **2022**, *194*, 826. [CrossRef]
- Li, X.; Yuan, W.; Dong, W. A Machine Learning Method for Predicting Vegetation Indices in China. *Remote Sens.* **2021**, *13*, 1147. [CrossRef]
- Sener, M.; Arslanoglu, M.C. Selection of the Most Suitable Sentinel-2 Bands and Vegetation Index for Crop Classification By Using Artificial Neural Network (Ann) and Google Earth Engine (Gee). *Fresenius Environ. Bull.* **2019**, *28*, 9348–9358.
- Marković, J.; Pavlović, M. *Geografske Regije Jugoslavije: (Srbija i Crna Gora)*; Savremena Administracija: Belgrade, Serbia, 1995.
- OpenStreetMap Contributors Planet OSM. 2017. Available online: <https://planet.osm.org> (accessed on 12 May 2023).
- European Environment Agency European Digital Elevation Model (EU-DEM)—Version 1.1; Copernicus Program. Available online: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata> (accessed on 30 December 2022).
- ESA Copernicus Open Access Hub Paris France, Hub. Available online: <https://scihub.copernicus.eu/> (accessed on 1 May 2023).

26. ESA User Guides—Sentinel-2 MSI—Sentinel Online—Sentinel Online. Available online: <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spectral> (accessed on 14 November 2022).
27. Dui, Z.; Huang, Y.; Jin, J.; Gu, Q. Automatic Detection of Photovoltaic Facilities from Sentinel-2 Observations by the Enhanced U-Net Method. *J. Appl. Remote Sens.* **2023**, *17*, 014516. [[CrossRef](#)]
28. Wang, H.; Zhang, L.; Wang, L.; He, J.; Luo, H. An Automated Snow Mapper Powered by Machine Learning. *Remote Sens.* **2021**, *13*, 4826. [[CrossRef](#)]
29. Stankevich, S.; Piestova, I.; Zaitseva, E.; Rusnak, P.; Rabcan, J. Satellite Imagery Spectral Bands Subpixel Equalization Based on Ground Classes' Topology. In Proceedings of the International Conference on Information and Digital Technologies 2019, Zilina, Slovakia, 25–27 June 2019.
30. Brodu, N. Super-Resolving Multiresolution Images with Band-Independent Geometry of Multispectral Pixels. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4610–4617. [[CrossRef](#)]
31. McGarragh, G.; Poulsen, C.; Povey, A.; Thomas, G.; Christensen, M.; Sus, O.; Schlundt, C.; Stapelberg, S.; Stengel, M.; Grainger, D.; et al. SNAP (Sentinel Application Platform) and the ESA Sentinel 3 Toolbox. *ESASP* **2015**, *734*, 21.
32. QGIS.org QGIS Geographic Information System. QGIS Association. Open Source Geospatial Foundation Project. 2022. Available online: <https://www.qgis.org/en/site/index.html> (accessed on 1 April 2023).
33. Silleos, N.G.; Alexandridis, T.K.; Gitas, I.Z.; Perakis, K. Vegetation Indices: Advances Made in Biomass Estimation and Vegetation Monitoring in the Last 30 Years. *Geocarto Int.* **2006**, *21*, 21–28. [[CrossRef](#)]
34. Kogan, F.N. Remote Sensing of Weather Impacts on Vegetation in Non-Homogeneous Areas. *Int. J. Remote Sens.* **1990**, *11*, 1405–1419. [[CrossRef](#)]
35. Liu, W.T.; Kogan, F.N. Monitoring Regional Drought Using the Vegetation Condition Index. *Int. J. Remote Sens.* **1996**, *17*, 2761–2782. [[CrossRef](#)]
36. Jackson, R.D.; Huete, A.R. Interpreting Vegetation Indices. *Prev. Vet. Med.* **1991**, *11*, 185–200. [[CrossRef](#)]
37. Richardson, J.F.; Wiegand, C.L. Distinguishing Vegetation from Soil Background Information (by Gray Mapping of Landsat MSS Data). *Photogramm. Eng. Remote Sens.* **1977**, *43*, 1541–1552.
38. Perry, C.R.; Lautenschlager, L.F. Functional Equivalence of Spectral Vegetation Indices. *Remote Sens. Environ.* **1984**, *14*, 169–182. [[CrossRef](#)]
39. Bannari, A.; Huete, A.R.; Morin, D.; Zagolski, F. Effets de La Couleur et de La Brilliance Du Sol Sur Les Indices de Végétation. *Int. J. Remote Sens.* **1996**, *17*, 1885–1906. [[CrossRef](#)]
40. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sens. Environ.* **1994**, *48*, 119–126. [[CrossRef](#)]
41. Ashburn, P.M. The Vegetative Index Number and Crop Identification. In *The LACIE Symposium Proceedings of the Technical Session*; NASA Johnson Space Center: Houston, TX, USA, 1979; Volume 1, pp. 843–850.
42. Huete, A.R.; Didan, K.; van Leeuwen, W.J.D.; Jacobson, A.; Solanos, R.; Laing, T.D. *Modis vegetation index (mod 13) algorithm theoretical basis document Version 3.1, Principal Investigators*; The University of Arizona: Tucson, AZ USA, 1999.
43. Pinty, B.; Verstraete, M.M. GEMI: A Non-Linear Index to Monitor Global Vegetation from Satellites. *Vegetatio* **1992**, *101*, 15–20. [[CrossRef](#)]
44. Gitelson, A.; Kaufman, J.; Merzlyak, N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
45. Clevers, J.G.P.W.; De Jong, S.M.; Epema, G.F.; van der Meer, F.; Bakker, W.H.; Skidmore, A.K.; Addink, E.A. Meris and the Red-Edge Index. In Proceedings of the 2nd EARSeL Workshop on Imaging Spectroscopy, Enschede, The Netherlands, 11–13 July 2000.
46. Guyot, G.; Baret, F. Utilisation de La Haute Resolution Spectrale Pour Suivre l'état Des Couverts Vegetaux. In Proceedings of the 4th International Colloquium on “Spectral Signatures of Objects in Remote Sensing”, Aussois, France, 18–22 January 1988; ESA SP-287. pp. 279–286.
47. Daughtry, C.S.T.; Walthall, C.L.; Kim, M.S.; De Colstoun, E.B.; McMurtrey, J.E. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sens. Environ.* **2000**, *74*, 229–239. [[CrossRef](#)]
48. Dash, J.; Curran, P.J. The MERIS Terrestrial Chlorophyll Index. *Int. J. Remote Sens.* **2004**, *25*, 5403–5413. [[CrossRef](#)]
49. Delegido, J.; Verrelst, J.; Alonso, L.; Moreno, J. Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content. *Sensors* **2011**, *11*, 7063–7081. [[CrossRef](#)] [[PubMed](#)]
50. Hardisky, M.A.; Klemas, V.; Smart, R.M. The Influence of Soil Salinity, Growth Form, and Leaf Moisture on the Spectral Radiance of *Spartina Alterniflora* Canopies. *Photogramm. Eng. Remote Sens.* **1983**, *49*, 77–83.
51. Cibula, W.G.; Zetka, E.F.; Rickman, D.L. Response of Thematic Mapper Bands to Plant Water Stress. *Int. J. Remote Sens.* **1992**, *13*, 1869–1880. [[CrossRef](#)]
52. Rouse, J.W.; Hass, R.H.; Schell, J.A.; Deering, D.W.; Harlan, J.C. *Monitoring the Vernal Advancement and Retrogradation (Green Wave Effect) of Natural Vegetation*; Final Report, RSC 1978-4; Texas A&M University: College Station, TX, USA, 1974.
53. Blackburn, G.A. Spectral Indices for Estimating Photosynthetic Pigment Concentrations: A Test Using Senescent Tree Leaves. *Int. J. Remote Sens.* **1998**, *19*, 657–675. [[CrossRef](#)]
54. Herrmann, I.; Pimstein, A.; Karnieli, A.; Cohen, Y.; Alchanatis, V.; Bonfil, D.J. LAI Assessment of Wheat and Potato Crops by VENÅμS and Sentinel-2 Bands. *Remote Sens. Environ.* **2011**, *115*, 2141–2151. [[CrossRef](#)]

55. Henrich, V.; Krauss, G.; Götze, C.; Sandow, C. *Entwicklung Einer Datenbank für Fernerkundungsindizes*; AK Fernerkundung: Bochum, Germany, 4–5 October 2012.
56. Gitelson, A.; Merzlyak, M.N. Quantitative Estimation of Chlorophyll-a Using Reflectance Spectra: Experiments with Autumn Chestnut and Maple Leaves. *J. Photochem. Photobiol. B Biol.* **1994**, *22*, 247–252. [[CrossRef](#)]
57. Birth, G.S.; McVey, G.R. Measuring the Color of Growing Turf with a Reflectance Spectrophotometer. *Agron. J.* **1968**, *60*, 640–643. [[CrossRef](#)]
58. Ahamed, T.; Tian, L.; Zhang, Y.; Ting, K.C. A Review of Remote Sensing Methods for Biomass Feedstock Production. *Biomass Bioenergy* **2011**, *35*, 2455–2469. [[CrossRef](#)]
59. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
60. Kauth, R.J. Tasselled cap—A graphic description of the spectral-temporal development of agricultural crops as seen by landsat. In Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, West Lafayette, IN, USA, 29 June–1 July 1976.
61. Deering, D.W.; Rouse, J.W.; Haas, R.H.; Schell, J.A. Measuring “forage production” of grazing units from landsat mss data. In Proceedings of the 10th International Symposium on Remote Sensing of Environment, Ann Arbor, MI, USA, 6–10 October 1975; Volume 2, pp. 1169–1178.
62. Gao, B.-C.; Yoram, J.K. *The MODIS Near-IR Water Vapor Algorithm: Product ID: MOD05—Total Precipitable Water*; Algorithm Technical Background Document; NASA: Washington, DC, USA, 1992.
63. Abdlaty, R.; Mokhtar, M. Toward Practical Analysis of Wastewater Contaminants Employing Dual Spectroscopic Techniques. *Water Conserv. Sci. Eng.* **2022**, *7*, 515–523. [[CrossRef](#)]
64. Li, Y.; Wu, Y.; Gao, Y.; Niu, X.; Li, J.; Tang, M.; Fu, C.; Qi, R.; Song, B.; Chen, H.; et al. Machine-Learning Based Prediction of Prognostic Risk Factors in Patients with Invasive Candidiasis Infection and Bacterial Bloodstream Infection: A Singled Centered Retrospective Study. *BMC Infect. Dis.* **2022**, *22*, 150. [[CrossRef](#)]
65. Shanbehzadeh, M.; Afrash, M.R.; Mirani, N.; Kazemi-Arpanahi, H. Comparing Machine Learning Algorithms to Predict 5-Year Survival in Patients with Chronic Myeloid Leukemia. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 236. [[CrossRef](#)]
66. Ramezan, C.A.; Warner, T.A.; Maxwell, A.E.; Price, B.S. Effects of Training Set Size on Supervised Machine-Learning Land-Cover Classification of Large-Area High-Resolution Remotely Sensed Data. *Remote Sens.* **2021**, *13*, 368. [[CrossRef](#)]
67. Rustam, F.; Reshi, A.A.; Mehmood, A.; Ullah, S.; On, B.W.; Aslam, W.; Choi, G.S. COVID-19 Future Forecasting Using Supervised Machine Learning Models. *IEEE Access* **2020**, *8*, 101489–101499. [[CrossRef](#)]
68. Armaghani, D.J.; Asteris, P.G.; Askarian, B.; Hasanipanah, M.; Tarinejad, R.; Huynh, V. Van Examining Hybrid and Single SVM Models with Different Kernels to Predict Rock Brittleness. *Sustainability* **2020**, *12*, 2229. [[CrossRef](#)]
69. Sanz, H.; Valim, C.; Vegas, E.; Oller, J.M.; Reverter, F. SVM-RFE: Selection and Visualisation of the Most Relevant Features through Non-Linear Kernels. *BMC Bioinform.* **2018**, *19*, 432. [[CrossRef](#)] [[PubMed](#)]
70. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
71. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
72. Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. A Practical Guide to Support Vector Classification. *BJU Int.* **2008**, *101*, 1–16.
73. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006; ISBN 9780387310732.
74. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
75. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in Neural Information Processing Systems*; Curran Associates: Red Hook, NY, USA, 2011; pp. 2951–2959.
76. WANG, D. Utilising Particle Swarm Optimisation to Optimise Hyper-Parameters of SVM Classifier. *J. Comput. Appl.* **2008**, *28*, 134–135. [[CrossRef](#)]
77. Wang, T.; Ye, X.; Wang, L.; Li, H. Grid Search Optimised SVM Method for Dish-like Underwater Robot Attitude Prediction. In Proceedings of the 2012 5th International Joint Conference on Computational Sciences and Optimization, Harbin, China, 23–26 June 2012.
78. Eskandari, A.; Milimonfared, J.; Aghaei, M. Optimization of SVM Classifier Using Grid Search Method for Line-Line Fault Detection of Photovoltaic Systems. In Proceedings of the Conference Record of the IEEE Photovoltaic Specialists Conference, Calgary, AB, Canada, 15 June–21 August 2020; Volume 2020. [[CrossRef](#)]
79. Bartz-Beielstein, T.; Zaefferer, M. Hyperparameter Tuning Approaches. In *Hyperparameter Tuning for Machine and Deep Learning with R*; Springer: Singapore, 2023; pp. 71–119.
80. Zhang, Q.; Fang, L.; Ma, L.; Zhao, Y. Research on Parameters Optimization of SVM Based on Improved Fruit Fly Optimisation Algorithm. *Int. J. Comput. Theory Eng.* **2016**, *8*, 500–505. [[CrossRef](#)]
81. Van Rossum, G.; Drake, F.L.; Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; et al. *Python 3 Reference Manual*; CreateSpace: Scotts Valley, CA, USA, 2009; Volume 585.
82. Hao, J.; Ho, T.K. Machine Learning Made Easy: A Review of Scikit-Learn Package in Python Programming Language. *J. Educ. Behav. Stat.* **2019**, *44*, 348–361. [[CrossRef](#)]
83. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 1–27. [[CrossRef](#)]
84. Gdal GDAL—Geospatial Data Abstraction Library. 2012. Available online: <https://gdal.org/> (accessed on 15 May 2023).
85. Gillies, S. Rasterio Documentation. Available online: <https://rasterio.readthedocs.io/en/stable/#n> (accessed on 15 May 2023).
86. Pyodbc. Available online: <https://pypi.org/project/pyodbc/> (accessed on 8 June 2023).

87. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)]
88. Esri Inc. *ArcGIS Pro, Version 3.0.3*; Esri Inc.: Redlands, CA, USA, 2023.
89. Microsoft Download Microsoft®SQL Server®2017 Express from Official Microsoft Download Center. Available online: <https://www.microsoft.com/en-us/download/details.aspx?id=55994> (accessed on 17 November 2022).
90. Stehman, S.V.; Fonte, C.C.; Foody, G.M.; See, L. Using Volunteered Geographic Information (VGI) in Design-Based Statistical Inference for Area Estimation and Accuracy Assessment of Land Cover. *Remote Sens. Environ.* **2018**, *212*, 47–59. [[CrossRef](#)]
91. Stehman, S.V.; Foody, G.M. Key Issues in Rigorous Accuracy Assessment of Land Cover Products. *Remote Sens. Environ.* **2019**, *231*, 111199. [[CrossRef](#)]
92. Fleuren, L.M.; Klausch, T.L.T.; Zwager, C.L.; Schoonmade, L.J.; Guo, T.; Roggeveen, L.F.; Swart, E.L.; Girbes, A.R.J.; Thorat, P.; Ercole, A.; et al. Machine Learning for the Prediction of Sepsis: A Systematic Review and Meta-Analysis of Diagnostic Test Accuracy. *Intensive Care Med.* **2020**, *46*, 383–400. [[CrossRef](#)] [[PubMed](#)]
93. Yang, M.; Xu, D.; Chen, S.; Li, H.; Shi, Z. Evaluation of Machine Learning Approaches to Predict Soil Organic Matter and PH Using Vis-NIR Spectra. *Sensors* **2019**, *19*, 263. [[CrossRef](#)] [[PubMed](#)]
94. Chen, R.C.; Dewi, C.; Huang, S.W.; Caraka, R.E. Selecting Critical Features for Data Classification Based on Machine Learning Methods. *J. Big Data* **2020**, *7*, 52. [[CrossRef](#)]
95. Coleman, C.; Kang, D.; Narayanan, D.; Nardi, L.; Zhao, T.; Zhang, J.; Bailis, P.; Olukotun, K.; Ré, C.; Zaharia, M. Analysis of Dawnbench, a Time-to-Accuracy Machine Learning Performance Benchmark. *Oper. Syst. Rev.* **2019**, *53*, 14–25. [[CrossRef](#)]
96. Madooei, A.; Abdlaty, R.M.; Doerwald-Munoz, L.; Hayward, J.; Drew, M.S.; Fang, Q.; Zerubia, J. Hyperspectral Image Processing for Detection and Grading of Skin Erythema. In Proceedings of the Medical Imaging 2017: Image Processing, Orlando, FL, USA, 11–16 February 2017; Volume 10133.
97. Shafique, A.; Cao, G.; Khan, Z.; Asad, M.; Aslam, M. Deep Learning-Based Change Detection in Remote Sensing Images: A Review. *Remote Sens.* **2022**, *14*, 871. [[CrossRef](#)]
98. Ming, Q.; Miao, L.; Zhou, Z.; Dong, Y. CFC-Net: A Critical Feature Capturing Network for Arbitrary-Oriented Object Detection in Remote-Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5605814. [[CrossRef](#)]
99. Li, W.; Dong, R.; Fu, H.; Yu, L. Large-Scale Oil Palm Tree Detection from High-Resolution Satellite Images Using Two-Stage Convolutional Neural Networks. *Remote Sens.* **2019**, *11*, 11. [[CrossRef](#)]
100. Baldeck, C.A.; Asner, G.P.; Martin, R.E.; Anderson, C.B.; Knapp, D.E.; Kellner, J.R.; Wright, S.J. Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. *PLoS ONE* **2015**, *10*, e0118403. [[CrossRef](#)]
101. Nasiri, V.; Sadeghi, S.M.M.; Moradi, F.; Afshari, S.; Deljouei, A.; Griess, V.C.; Maftai, C.; Borz, S.A. The Influence of Data Density and Integration on Forest Canopy Cover Mapping Using Sentinel-1 and Sentinel-2 Time Series in Mediterranean Oak Forests. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 423. [[CrossRef](#)]
102. Nasiri, V.; Deljouei, A.; Moradi, F.; Sadeghi, S.M.M.; Borz, S.A. Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods. *Remote Sens.* **2022**, *14*, 1977. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.