

## Article

# Water Hyacinth Invasion and Management in a Tropical Hydroelectric Reservoir: Insights from Random Forest and SVM Classification

Luis Fernando Correa-Mejía and Yeison Alberto Garcés-Gómez \* 

Faculty of Engineering and Architecture, Universidad Católica de Manizales, Manizales 170002, Colombia; luis.correa1@ucm.edu.co

\* Correspondence: ygarces@ucm.edu.co

**Abstract:** The rapid proliferation of water hyacinth (*Eichhornia crassipes*) in newly formed reservoirs poses a significant threat to aquatic ecosystems and hydroelectric operations. The objective of this study was to map and monitor the spatio-temporal distribution of water hyacinth in the Hidroituango reservoir in Colombia from 2018 to 2023, using Sentinel-2 satellite imagery and machine learning algorithms. The Random Forest (RF) and Support Vector Machine (SVM) algorithms were employed for image classification, and their performance was evaluated using various accuracy metrics. The results revealed that both algorithms effectively detected and mapped water hyacinth infestations, with RF demonstrating greater stability in capturing long-term trends and SVM exhibiting higher sensitivity to rapid changes in coverage. The study also highlighted the impact of the COVID-19 pandemic on control efforts, leading to a temporary increase in infestation. The findings underscore the importance of continuous monitoring and adaptive management strategies to mitigate the ecological and economic impacts of water hyacinth in the Hidroituango reservoir and similar environments.

**Keywords:** Sentinel-2; *Eichhornia crassipes*; machine learning; image classification; ecological impact; economic impact



Academic Editor: José Ramón Arévalo Sierra

Received: 3 September 2024

Revised: 20 January 2025

Accepted: 21 January 2025

Published: 23 January 2025

**Citation:** Correa-Mejía, L.F.; Garcés-Gómez, Y.A. Water Hyacinth Invasion and Management in a Tropical Hydroelectric Reservoir: Insights from Random Forest and SVM Classification. *Ecologies* **2025**, *6*, 8. <https://doi.org/10.3390/ecologies6010008>

**Copyright:** © 2025 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/>).

## 1. Introduction

The Ituango Hydroelectric Project (Hidroituango), whose construction began in 2010, is situated in the northern region of the department of Antioquia, approximately 98 km NNW of the city of Medellín. On 30 April 2018, when the dam's height reached 175 m, an unplanned filling of the reservoir occurred due to a blockage in the Auxiliary Diversion Tunnel (ADT). Following several months of contingency measures, the spillway was partially operationalized on 4 November 2018, starting to discharge at an elevation of 404 m.a.s.l., which allowed for the regulation and maintenance of a relatively uniform reservoir level [1].

A few months after the formation of the reservoir, the first appearances of water hyacinth (*Eichhornia crassipes*) were reported on its surface. To monitor the expansion of this invasive species and assess its impact on the ecosystem and communities, satellite imagery was utilized. Analysis of these images revealed the magnitude of the invasion, showing the obstruction of navigation routes and the impact on ferry services. Remote sensing became a crucial tool to assess the rapid proliferation of this plant in a dynamic environment, highlighting the importance of early detection and accurate mapping through satellite imagery. By early 2019, the colonies had caused problems with navigability, hindering the

normal transit of small boats between the dam site and the tail of the reservoir, impacting ferry services for local communities, and leading to complaints from fishermen living around the reservoir.

Water hyacinth (*Eichhornia crassipes* (Martius) Solms-Laubach) is a perennial, free-floating, monocotyledonous aquatic macrophyte belonging to the Pontederiaceae family. It has lilac flowers with variations from blue to purple, featuring a yellow spot on one of its six segments. The leaves are circular and bulbous, typically less than 30 cm in size. Since the 19th century, it has spread anthropogenically across the globe from its native range in the tropical regions of South America, specifically the Amazon basin [2]. Water hyacinth forms dense mats on the water surface, blocking light penetration for native aquatic plants and reducing dissolved oxygen levels [3]. Its reproduction is sexual, occurring in seasonal habitats with frequent water level fluctuations, which provide suitable conditions for seed germination and seedling establishment [4].

The seeds can germinate within a few days or remain submerged and dormant for 15 to 20 years. They typically sink and remain dormant during dry periods. Upon re-flooding, the seeds often germinate and renew the growth cycle. Although not fully understood, natural populations can flower repeatedly throughout the year under favorable growth conditions, though flowering intensity may vary with seasonal changes and growth rates [3].

Water hyacinth thrives in a wide range of tropical and subtropical wetlands, preferring nutrient-rich waters, though it can tolerate considerable variations in nutrients, temperature, and pH. Its optimal pH for growth is between 6 and 8. It can grow in temperatures ranging from 1 °C to 40 °C, with optimal growth between 25 °C and 27.5 °C, being sensitive to cold [5]. High salt concentrations (6–8‰) are lethal and inhibit its growth [3]. This plant reproduces rapidly, covering large areas of water, reducing the amount of light and oxygen available for aquatic organisms, affecting water quality and ecosystem biodiversity [6]. Additionally, it can clog gates and turbines of hydroelectric power plants, causing significant economic damage [7].

Over the past three decades, the literature has shown a significant increase in studies employing satellite sensors to estimate water hyacinth infestations [8]. Most of these studies focus on large bodies of water such as lakes and reservoirs, with little attention given to small rivers [2]. The availability of satellite sensors like Sentinel-2, with high revisit frequency, broad coverage, and high radiometric, spatial, and spectral resolution, has facilitated new studies adopting remote sensing technologies for aquatic weed management [9–11].

Satellite data offers an effective, cost-efficient, and frequent means of monitoring the spatial and temporal distribution of these infestations on a large scale [2]. Multispectral satellite imagery enables natural resource managers and hydroelectric project operators to detect the presence of the plant, assess its extent and density, predict its spread, and plan control and eradication strategies [12]. Machine learning algorithms, such as Random Forest (RF) and Support Vector Machine (SVM), are widely used in remote sensing image analysis due to their accuracy and robustness in classifying high-dimensional data [13]. Specifically, studies have shown high accuracy in detecting and mapping water hyacinth using RF and SVM algorithms, achieving up to 85% accuracy with RF compared to 65% with SVM [14].

This study aimed to map and monitor the spatio-temporal distribution of water hyacinth in the Hidroituango reservoir in Colombia from 2018 to 2023, using Sentinel-2 satellite imagery and machine learning algorithms.

## 2. Materials and Methods

### 2.1. Study Area

The Hidroituango reservoir, a prominent hydroelectric infrastructure project, is in Colombia, specifically in the department of Antioquia, in the northwestern region of the country. Situated on the Cauca River, one of Colombia's most significant waterways, it forms an integral part of the Ituango Hydroelectric Project. Geographically, the reservoir lies in the middle section of the Cauca River canyon. This canyon, which separates the Central and Western Cordilleras of the Colombian Andes, constitutes its natural surroundings (See Figure 1). With an approximate flooding area of 31 km<sup>2</sup>, it has an average elevation of around 420 m.a.s.l. The average water temperature of the Cauca River near the tail of the reservoir is 26.21 °C. During the dry season, the water has an average temperature of 27.50 °C and one of 25.28 °C in the wet season. Near the dam site, the average temperature is 26.48 °C. During the dry season, the water has an average temperature of 26.57 °C and one of 26.03 °C in the wet season. The existing hydrological regime of the Cauca River is characterized by the bimodality present in the Colombian climate, represented by two dry seasons: December–January and July–August, and two rainy seasons: April–May and October–November [15]. The reservoir has a maximum operating level of 420 m.a.s.l. and a minimum operating level of 390 m.a.s.l.

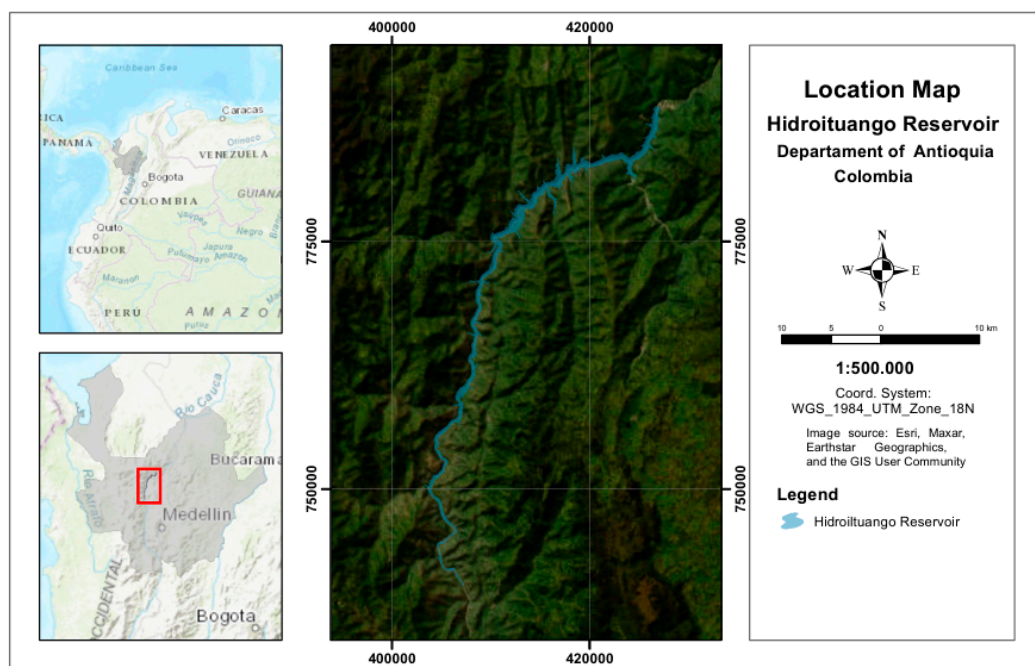


**Figure 1.** General location map of Hidroituango reservoir. Department of Antioquia, Colombia.

The annual average temperature in the Hidroituango reservoir ranges from 24 °C to 27 °C, with the pH between 6.8 and 9. Nitrates range from 0.2 mgN/L to 1.1 mgN/L, phosphates between 0.0074 mgP/L and 6.8 mgP/L, and flow velocities between 0 m/s and 1.7 m/s. These annual average values indicate a high availability of nutrients and optimal temperature and pH conditions for the growth of the species [15].

Empresas Públicas de Medellín (EPM), as the owner of the Ituango Hydroelectric Project, has committed to the National Environmental Licensing Authority (ANLA) and the environmental license to maintain water hyacinth coverage in the reservoir below 20%, with a maximum limit of 600 hectares, and to properly manage the removed material. To control the macrophytes, EPM carries out the extraction of plant material using mechanical methods, an ongoing task since September 2018, at a rate of 1800 m<sup>3</sup> per day. By November

2020, it is estimated that approximately 490,000 m<sup>3</sup> of material had been extracted. These extraction activities have successfully kept the spread of water hyacinth in the reservoir under control.

## 2.2. Dataset and Preprocessing

After acquiring Sentinel-2 Level 2A images spanning a period of six years, preprocessing of these images is conducted. Subsequently, automatic detection of water hyacinth is carried out using supervised learning methods, employing two algorithms: Random Forest (RF) and Support Vector Machine (SVM). The classification results are utilized in the mapping phase, providing the percentage of water hyacinth detected over time within the reservoir. Figure 2 illustrates the workflow employed for mapping water hyacinth throughout the analyzed time series.

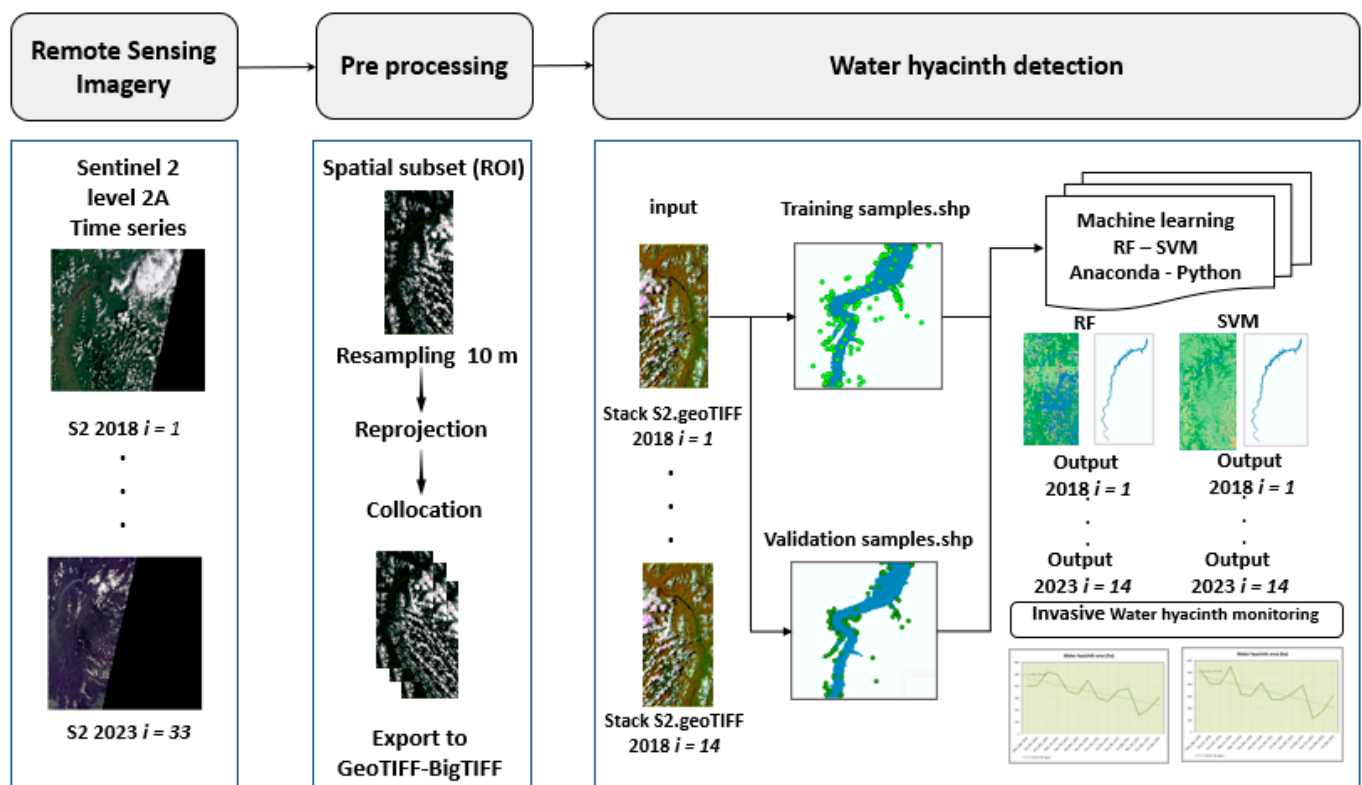


**Figure 2.** A methodology for mapping the spatio-temporal distribution of water hyacinth in the Hidroituango reservoir using the RF and SVM algorithms.

Among the advanced resources for studying the Earth's surface, the satellites launched by the European Space Agency (ESA) as part of the Copernicus program stand out. Among these, the Sentinel-2 (S2) mission emerges as a valuable resource by providing freely accessible multispectral datasets. Level-2A (S2L2A) products, which include atmospheric correction, were used, obtained through the SentinelHub platform. These S2L2A images provide a versatile set of 13 spectral bands: 4 visible bands, 6 near-infrared bands, and 3 shortwave infrared bands with a revisit time of 5 days, and with a spatial resolution of 10 m, 20 m, and 60 m depending on the wavelength, providing a dynamic and detailed view of changes in water and vegetation cover. Table 1 provides details on the spatial and spectral resolutions of these images in each of the 13 available spectral bands. For the study, a 6-year period was monitored using a total of 33 S2 images, collected between May 2018 and August 2023, excluding those affected by the presence of clouds in the region of interest (ROI). Table 2 presents the final set of images used in the research.

**Table 1.** Spatial and spectral resolutions of Sentinel 2 image bands.

| Band              | Spatial Resolution (m) | Wavelength (nm) |
|-------------------|------------------------|-----------------|
| B2 (blue)         | 10                     | 496.6           |
| B3 (green)        | 10                     | 560             |
| B4 (red)          | 10                     | 664.5           |
| B5 (red edge 1)   | 20                     | 703.9           |
| B6 (red edge 2)   | 20                     | 740.2           |
| B7 (red edge 3)   | 20                     | 782.5           |
| B8 (NIR)          | 10                     | 835.1           |
| B8A (NIR narrow)  | 20                     | 864.8           |
| B11 (SWIR cirrus) | 20                     | 1613.7          |
| B12 (SWIR)        | 20                     | 2202.4          |

**Table 2.** Number of Sentinel 2 images processed per month and year.

| Year         | Jan      | Feb      | Mar      | Apr      | May      | Jun      | Jul      | Aug      | Sep      | Oct      | Nov      | Dec      | Total     |
|--------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|
| 2018         |          |          |          |          | 1        |          |          | 1        | 1        |          |          | 1        | 4         |
| 2019         |          | 1        | 1        |          | 1        | 1        | 1        | 1        |          | 2        |          |          | 8         |
| 2020         |          | 1        | 1        |          |          |          |          |          | 1        |          | 1        |          | 4         |
| 2021         |          | 1        |          | 1        |          |          | 1        | 1        |          |          | 1        |          | 5         |
| 2022         | 1        | 1        |          | 1        |          |          | 1        | 1        |          |          | 1        | 1        | 7         |
| 2023         | 1        |          | 1        |          |          |          | 2        | 1        |          |          |          |          | 5         |
| <b>Total</b> | <b>2</b> | <b>4</b> | <b>3</b> | <b>2</b> | <b>2</b> | <b>1</b> | <b>5</b> | <b>5</b> | <b>2</b> | <b>2</b> | <b>3</b> | <b>2</b> | <b>33</b> |

As mentioned, Sentinel 2 Level 2A images were used, which are orthorectified and have reflectance levels below the atmosphere (BOA). Images with a cloud cover percentage of less than 5% over the reservoir area were selected. Preprocessing was performed using the public domain SNAP software (Sentinel Application Platform), a tool developed by the European Space Agency (ESA) for processing, analyzing, and visualizing Sentinel satellite data. The following steps were followed:

- **Resampling:** The spatial resolution of the bands was modified, bringing them all to 10 m using one of the visible bands as a reference. In this process, each pixel of the resampled bands takes the value of the nearest pixel in the original resolution, without interpolation. This preserves the integrity of the categorical data. An advantage of the nearest neighbor method is its ability to conserve the original data values, avoiding distortions that could affect the analysis of discrete data, such as classifications, and it is computationally efficient [16]. The result is a set of bands homogeneous in spatial resolution, suitable for multiband analysis, such as creating spectral indices.
- **Subset:** The images were cropped to a region of interest (ROI) that included the entire reservoir to facilitate processing. Additionally, a subset was created with bands B2, B3, B4, B5, B6, B7, B8, B8A, B9, B11, and B12.
- **Collocation:** Finally, an image stacking was performed, using a maximum of three images per period through the Collocation tool. This tool allows for the spatial overlay of images so that the pixel values of the secondary products (dependent images) are sampled on the geographic grid of the main product (reference image).

### 2.3. Training and Validation Points

The use of machine learning methods for water hyacinth mapping, employing a supervised classification approach, necessitates training samples as input for classification.

The images were categorized into five land cover classes: Water, Water Hyacinth, Other Vegetation, Bare Soil, and Grasses. The Water class encompassed the reservoir's water body, regardless of its turbidity or depth. The Water Hyacinth class was assigned

to areas with the presence of this macrophyte, irrespective of its phenological state. It is important to note that the presence of other macrophyte species besides water hyacinth is possible in the identified infestation areas, and their categorization was not the focus of this study. The Other Vegetation category included forested and shrub areas located on the reservoir's shores. On the other hand, the Bare Soil class included areas with sediment deposition, alluvial bars, areas affected by landslides on the reservoir margins, and rocky outcrops exposed during periods of low water levels. The Grasses class encompassed areas where rapid growth of this vegetation was observed on sediment bars. It is worth noting that the development of grasses on mats generated by water hyacinth is common.

The training and validation sets were constructed through a visual interpretation process, based on a meticulous analysis of a high-resolution orthophoto from 2021, as well as high-resolution images obtained from Google Earth and Bing Maps Satellite View. Additionally, some field data were collected for the year 2023 using a GPS navigator, with a particular focus on water hyacinth infestation areas and on grass areas developed on sediment bars.

The creation of the training and validation point sets was carried out using QGIS software, where each point was appropriately labeled.

#### 2.4. Image Classification with Machine Learning

During each phase of the classification process, specific development and programming environments were employed. In the initial stage, Anaconda, a Python distribution that integrates a wide range of packages and tools essential for data analysis and development in this language, was utilized [17]. For processing, the following libraries were used: NumPy 2.0.0, Matplotlib 3.7.2, Matplotlib.pyplot 3.7.2, Pandas 2.0.3, GeoPandas 0.14.0, Snappy 1.1.10, EarthPy 0.9.4, and Rasterio 1.3.0.

Furthermore, Jupyter Notebook, an open-source web application that facilitates the combination of executable code with detailed explanations, graphics, and results, was utilized. This tool enabled an interactive presentation of the processes and results obtained.

For the image classification, two non-parametric machine learning methods were applied: the Random Forest (RF) [13,18–20] and Support Vector Machine [21] algorithms. The RF classifier was trained to differentiate between classes in the model based on the dB values of the bands in the previously created training data. A number of trees in the range of 50 to 500 has been shown to perform well, with greater accuracy for RF classification. Bayable et al. (2023) found that a number of 300 trees produces good results in the analysis with multispectral images; therefore, the number of trees was set to 300 for the present study [22]. Table 3 details the parameters and default values used in the RF classification.

**Table 3.** Parameters and values used in the RF classifier.

| Parameter                | Value | Description                                                                                                                          |
|--------------------------|-------|--------------------------------------------------------------------------------------------------------------------------------------|
| max_depth                | None  | The nodes expand until all sheets are pure or contain fewer samples than the value defined in min_samples_split.                     |
| min_samples_split        | 2     | The minimum number of samples required to split an internal node. Controls the creation of new splits and helps prevent overfitting. |
| min_samples_leaf         | 1     | It represents the minimum number of samples required to be in a leaf node.                                                           |
| min_weight_fraction_leaf | 0     | It represents the minimum weighted fraction of the total sum of sample weights required to be in a leaf node.                        |
| max_features             | auto  | All characteristics are considered when looking for the best division.                                                               |
| max_leaf_nodes           | None  | There is no limit to the number of leaf nodes allowed in each tree.                                                                  |
| min_impurity_decrease    | 0     | Indicates that a node will split if this division induces a decrease in impurity greater than or equal to this value.                |

**Table 3.** *Cont.*

| Parameter         | Value | Description                                                                                                                      |
|-------------------|-------|----------------------------------------------------------------------------------------------------------------------------------|
| bootstrap         | True  | Sampling with replacement was used when building trees.                                                                          |
| oob_score         | False | It indicates that the score is not calculated outside the bag to evaluate the generalization of the model.                       |
| n_jobs            | None  | Indicates that only one job is used during tuning and prediction.                                                                |
| random_state      | None  | It uses the global random seed from the Numpy library.                                                                           |
| verbose:          | 0     | No output during adjustment.                                                                                                     |
| warm_start        | False | Previous results are not reused to adjust and add additional trees to the estimator.                                             |
| class_weight      | None  | All classes have equal weight in the classification problem.                                                                     |
| ccp_alpha         | 0     | Represents the pruning complexity parameter.                                                                                     |
| max_samples       | None  | Use all the samples when adjusting each tree.                                                                                    |
| max_depth         | None  | The nodes expand until all sheets are pure or contain fewer samples than the value defined in min_samples_split.                 |
| min_samples_split | 2     | Minimum number of samples required to split an internal node. Controls the creation of new splits and helps prevent overfitting. |

Table 4 details the parameters and default values used in the SVM classification.

**Table 4.** Parameters and values used in the SVM classifier.

| Parameter               | Value | Description                                                                                                |
|-------------------------|-------|------------------------------------------------------------------------------------------------------------|
| C                       | 1     | Control the trade-off between the soft decision limit and the correct ranking of training points.          |
| kernel                  | rbf   | Radial kernel suitable for most cases.                                                                     |
| degree                  | 3     | Polynomial kernel grade.                                                                                   |
| gamma                   | scale | Uses $1/(n\_features * X.var())$ as the gamma value.                                                       |
| coef0                   | 0     | A standalone term in the kernel function.                                                                  |
| shrinking               | True  | Uses the support vector reduction heuristic.                                                               |
| probability             | False | Indicates whether probability estimates should be enabled.                                                 |
| tol                     | 0.001 | Tolerance to stop the adjustment process. Controls the accuracy of the result.                             |
| cache_size              | 200   | Cache size in MB.                                                                                          |
| class_weight            | None  | All classes have the same weight.                                                                          |
| verbose                 | False | Controls the verbosity of the output during tuning.                                                        |
| max_iter                | −1    | Maximum number of iterations. There is no limit.                                                           |
| decision_function_shape | ovr   | Specifies the shape of the decision function. ‘OVR’ (one against the rest).                                |
| break_ties:             | True  | If there is a tie in the class decision, the tie will be broken according to the score of the classifiers. |
| random_state            | None  | Controls the randomness of the estimator.                                                                  |

### 2.5. Assessing Accuracy

After completing the classification of the 2021 image series using both algorithms, the model’s performance was evaluated through a series of parameters that offer a detailed understanding of its performance. These parameters include the following:

- **Precision:** This measures the proportion of instances classified as positive that are actually positive (TP) with respect to instances classified as false positives (FP). In other words, how accurate the model’s positive predictions are.

$$Precision = \frac{TP}{TP + FP}$$

- Recall: This indicates the proportion of positive instances that were correctly identified by the model with respect to instances classified as false negatives (FN). It is especially important in problems where the omission of positive instances is critical.

$$Recall = \frac{TP}{TP + FN}$$

- F1-score: This is a measure that combines Precision and Recall into a single value, calculating the harmonic mean between them. It provides a balance between the model's precision and recall capabilities.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

These three metrics are key indicators for measuring the quality of the classification in terms of the model's ability to make accurate and complete predictions for each class [23].

In addition to these metrics, other parameters were calculated to evaluate the overall performance of the model:

- Accuracy: This provides a global measure of the proportion of correct predictions made by the model over the total number of predictions made.

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

- Macro average (Macro avg): This offers a broader evaluation of the model's performance by considering the class balance in the dataset. The macro average calculates the unweighted average of the metrics for each class.
- Weighted average (Weighted avg): This also considers the class balance but considers the weight of each class based on its support in the dataset.

These additional parameters provide valuable information about how the model performs under different conditions and contexts, allowing for a more comprehensive evaluation of its predictive capability.

### 3. Results and Discussion

This section describes the results of water hyacinth detection obtained from the two classifiers used. Table 5 presents the results of the evaluated metrics for the training set. A set of 1280 points was generated for the training data, with approximately 200 points per class, and an independent set of 588 points for the validation data, as follows: Water (162 points), Water Hyacinth (136 points), Other Vegetation (157 points), Bare Soil (39 points), and Grasses (94 points). It is important to highlight that the Bare Soil and Grasses classes had the fewest training and validation points since they are found in small patches and occupy the smallest area within the target classes during the study period.

Based on the results of precision, recall, and F1-score for the Random Forest (RF) and Support Vector Machine (SVM) classifiers, the following conclusions can be drawn regarding each class:

- Water: both classifiers exhibit very similar and highly accurate results for the Water class.
- Water Hyacinth: while both techniques show good performance, RF appears to be slightly more precise in classifying Water Hyacinth.
- Other Vegetation: both classifiers provide consistent and high results for the Other Vegetation class.

- Bare Ground: both classifiers have high precision but lower recall for the Bare Soil class, indicating that there may be some instances of bare soil that were not correctly classified, particularly with SVM.
- Grasses: both classifiers offer similar and high results for the Grasses class.

**Table 5.** Metrics evaluated for RF and SVM algorithm training set.

| Class                 | RF          |             |            |            | SVM         |             |             |            |
|-----------------------|-------------|-------------|------------|------------|-------------|-------------|-------------|------------|
|                       | Precision   | Recall      | F1-Score   | Support    | Precision   | Recall      | F1-Score    | Support    |
| Water                 | 0.98        | 0.95        | 0.97       | 162        | 0.97        | 0.95        | 0.96        | 162        |
| <b>Water Hyacinth</b> | <b>0.86</b> | <b>0.96</b> | <b>0.9</b> | <b>136</b> | <b>0.81</b> | <b>0.93</b> | <b>0.87</b> | <b>136</b> |
| Other Vegetation      | 0.9         | 0.92        | 0.91       | 157        | 0.87        | 0.92        | 0.9         | 157        |
| Bare Ground           | 1           | 0.72        | 0.84       | 39         | 1           | 0.62        | 0.76        | 39         |
| Grass                 | 0.92        | 0.88        | 0.9        | 94         | 0.94        | 0.85        | 0.89        | 94         |
| Accuracy              |             |             | 0.92       | 588        |             |             | 0.9         | 588        |
| Macro Avg             | 0.93        | 0.89        | 0.9        | 588        | 0.92        | 0.85        | 0.88        | 588        |
| Weighted Avg          | 0.9         | 0.92        | 0.92       | 588        | 0.91        | 0.9         | 0.9         | 588        |

Overall, these findings suggest that both RF and SVM are effective in classifying the different land cover classes in the study area, with RF potentially having a slight edge in distinguishing Water Hyacinth and SVM potentially having a slight disadvantage in identifying Bare Soil. The high accuracy and consistency across most classes demonstrate the suitability of these machine learning techniques for mapping water hyacinth infestations in the Hidroituango reservoir.

Figure 3 shows the dynamics exhibited by water hyacinth in one of the critical infestation hotspots in the reservoir. Specifically, it reports the results obtained after applying the RF algorithm in each of the periods from 2018 to 2023. On the other hand, Figure 4 shows the results of this hotspot obtained using the SVM algorithm.

Once the mapping of the evaluated land covers was obtained, the areas in hectares and percentages were estimated according to the extent of the reservoir in each period. Figures 5 and 6 present the graphs of the area in percentage by cover type.

To analyze the results of water hyacinth coverage in the Hidroituango reservoir, it is crucial to remember that the reservoir was created in April 2018, and within a few months of its establishment, extensive colonies were detected, hindering the normal movement of boats and ferries. In September 2018, control measures began through the mechanical removal of hyacinth mats and the placement of floating barriers to prevent its spread. No published information was found on the volumes removed in each period. Furthermore, the analyzed years cover the pandemic and post-pandemic period, during which hyacinth control was possibly irregular.

During the period from May to August 2018, water hyacinth covered an area corresponding to 13.06% according to the RF classifier and 17.03% according to the SVM classifier of the total reservoir area. Assuming that the first hyacinth seeds arrived with the formation of the reservoir and control began in September 2018, both algorithms show an approximate growth rate of water hyacinth of 25 hectares per month. This indicates that, based on the obtained data, both algorithms agree on the estimated growth rate during this initial period before the start of control in September 2018. This figure is consistent with the high growth rates reported under optimal temperature and nutrient conditions [7,24].

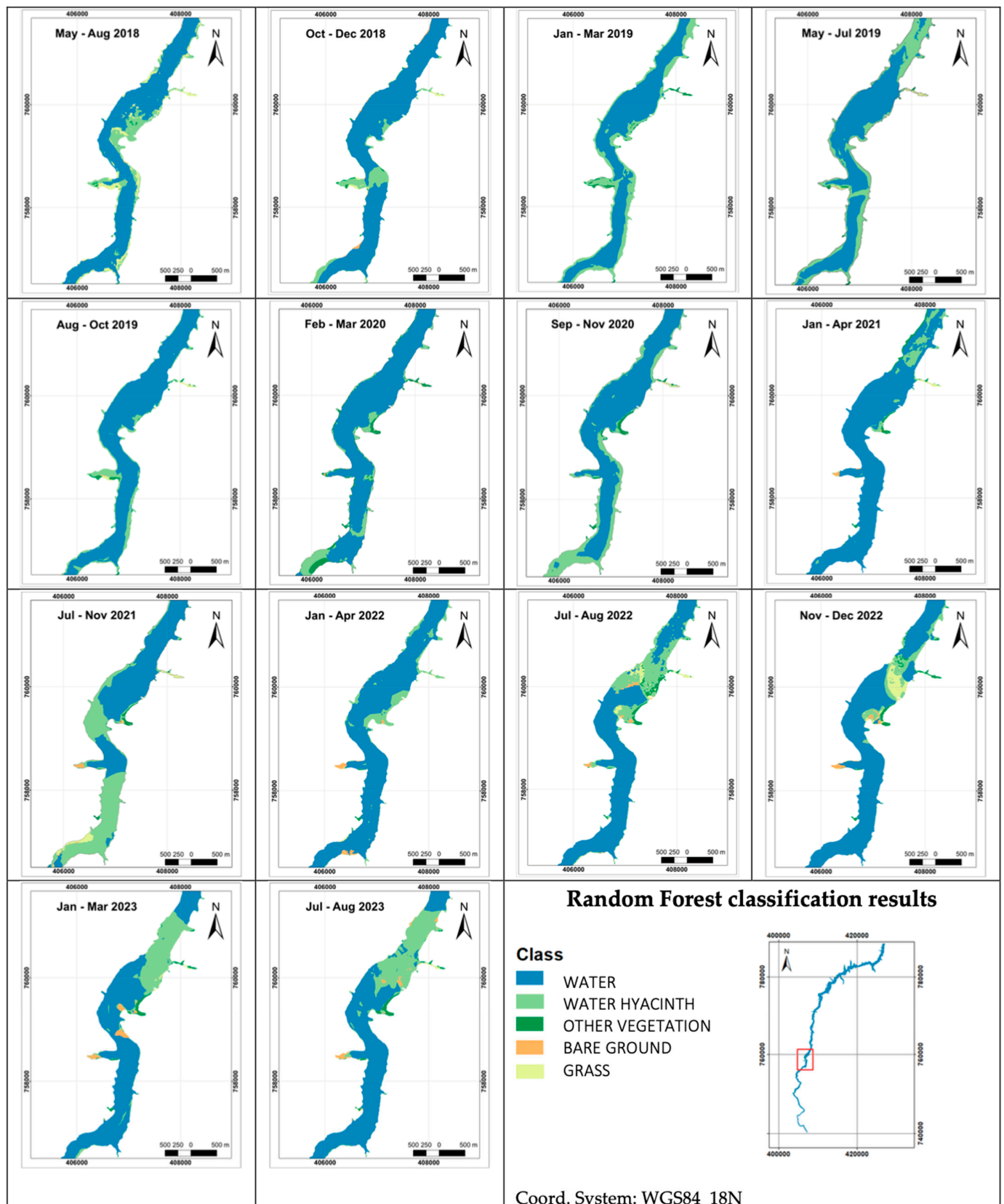


**Figure 3.** Mapping results obtained after applying the RF algorithm to the Sentinel 2 images, in each of the periods evaluated between 2018 and 2023. The area shown corresponds to a hotspot of reservoir infestation during all seasons.

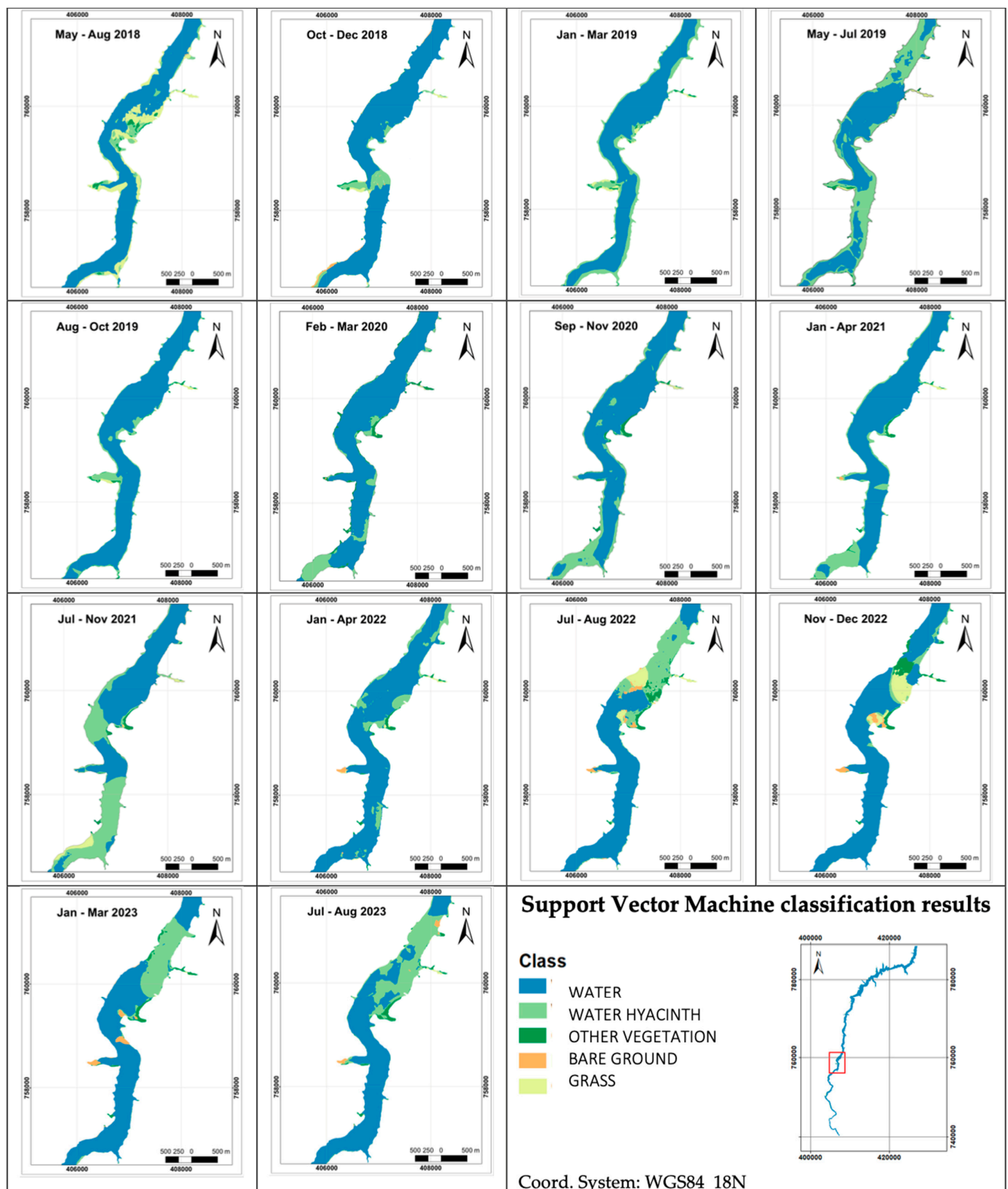


**Figure 4.** Mapping results obtained after applying the SVM algorithm to the Sentinel 2 images, in each of the periods evaluated between 2018 and 2023. The area shown corresponds to a hotspot of reservoir infestation during all seasons.

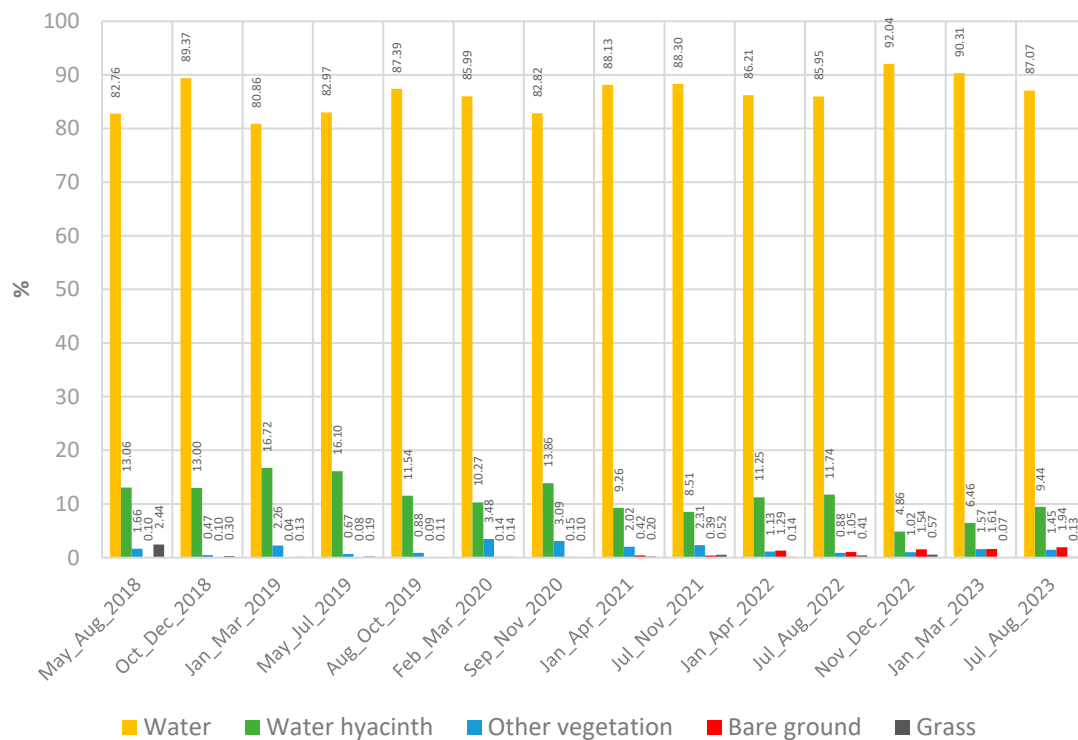


**Figure 5.** Percentage of area estimated by RF for each coverage within the reservoir.

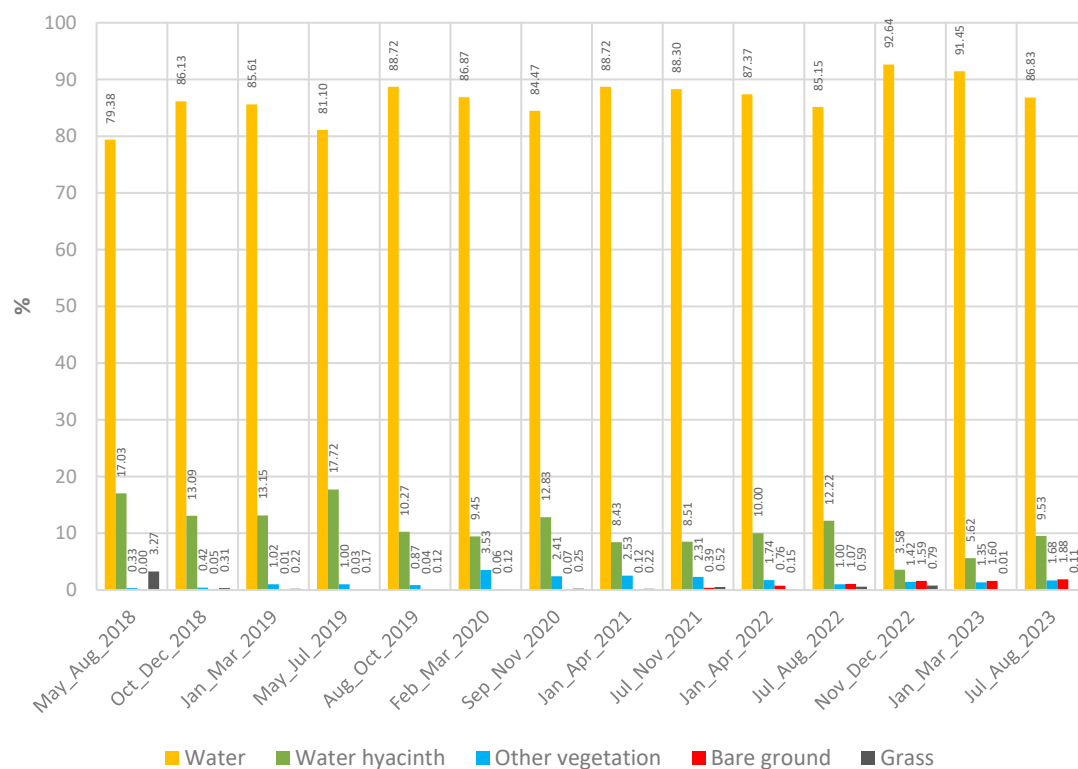


**Figure 6.** Percentage of area estimated by SVM for each coverage within the reservoir.

It is important to highlight that logistic growth models describe how a population can experience rapid growth initially and then stabilize as it approaches the carrying capacity of the environment. This behavior is due to resource limitations and intraspecific competition, which eventually slow down the growth rate as the population nears its sustainable limit. In this context, the observed growth rate of water hyacinth can be influenced by various

environmental factors, including nutrient availability and water temperature, which are crucial for its proliferation [5].

This study presents a novel approach to understanding water hyacinth dynamics in the Hidroituango reservoir, a newly formed ecosystem with a unique history due to an unplanned filling event. By utilizing a six-year dataset, the research captures the impact of the COVID-19 pandemic on water hyacinth control efforts, revealing the consequences of disruptions in management activities and the subsequent recovery.

The research distinguishes itself through a detailed comparison of Random Forest (RF) and Support Vector Machine (SVM) algorithms for water hyacinth mapping. This comparative analysis highlights the strengths and weaknesses of each method, with RF demonstrating stability for long-term trend analysis and SVM exhibiting sensitivity to rapid changes for immediate control response planning. The study proposes combining both algorithms to leverage their complementary strengths for a comprehensive and adaptive management approach.

Furthermore, the study provides a thorough methodological workflow, encompassing image preprocessing, training and validation data collection, and classification procedures. This detailed description ensures the reproducibility of the research and offers a valuable guide for similar studies in other regions.

The study successfully mapped and monitored the spatio-temporal distribution of water hyacinth in the Hidroituango reservoir from 2018 to 2023 using Sentinel-2 satellite imagery and machine learning algorithms. Both Random Forest (RF) and Support Vector Machine (SVM) algorithms effectively detected and mapped water hyacinth infestations. However, RF demonstrated greater stability in capturing long-term trends, while SVM exhibited higher sensitivity to rapid changes in coverage. These results are similar to those reported in the literature for remote sensing applications [25].

The study also revealed the impact of the COVID-19 pandemic on control efforts, leading to a temporary increase in infestation. This highlights the importance of continuous monitoring and adaptive management strategies for mitigating the ecological and economic impacts of water hyacinth in the reservoir and similar environments.

The findings contribute valuable insights for managing water hyacinth in newly formed reservoirs and underscore the importance of integrating remote sensing and machine learning for effective monitoring and control of invasive species. Future research could explore integrating high-resolution imagery, investigating alternative machine learning algorithms, developing early warning systems, assessing climate change impacts, and evaluating diverse control strategies to enhance the understanding and management of water hyacinth in the Hidroituango reservoir and similar ecosystems.

#### 4. Conclusions

The results obtained through the Random Forest (RF) and Support Vector Machine (SVM) algorithms for detecting and monitoring water hyacinth in the Hidroituango reservoir reveal significant conclusions about the effectiveness of classification methods and the dynamics of hyacinth infestation and control throughout the evaluated period from 2018 to 2023.

- Algorithm performance:
  - Random Forest: provided a smoother and more stable view of water hyacinth dynamics, beneficial for analyzing long-term trends and gradual changes.
  - Support Vector Machine: Indicated greater sensitivity to fluctuations in water hyacinth coverage, capturing more pronounced variations and reflecting higher sensitivity to changes in control conditions. This makes it suitable for detecting rapid changes and peaks in infestation.

- Impact of the COVID-19 pandemic:
  - During the COVID-19 pandemic (March 2020–July 2022), control activities were intermittent or suspended, leading to a significant increase in the area covered by water hyacinth according to both algorithms. This highlights the importance of maintaining infestation control and the need for continued implementation of control measures.
- Post-pandemic control and recovery:
  - After the resumption of post-pandemic control activities, a decrease in the area occupied by water hyacinth is observed in both algorithms, indicating the effectiveness of the implemented measures.
- Algorithm strengths and applications:
  - RF's stability advantage: allows for long-term monitoring and trend analysis, providing a reliable tool for assessing the effectiveness of control policies over time.
  - SVM's sensitivity to rapid changes: makes it beneficial for reactive management and emergency control measure planning due to its high sensitivity to data variability.

In conclusion, both RF and SVM proved valuable in understanding the dynamics of water hyacinth infestation and control in the Hidroituango reservoir. The choice between them depends on the specific management objectives. RF is ideal for long-term monitoring and trend analysis, while SVM is better suited for detecting rapid changes and planning immediate control responses. Combining both algorithms could offer a comprehensive approach to water hyacinth management, leveraging their complementary strengths for effective and adaptive control strategies.

**Author Contributions:** Conceptualization, L.F.C.-M. and Y.A.G.-G.; methodology, Y.A.G.-G.; software, L.F.C.-M.; validation, Y.A.G.-G.; formal analysis, L.F.C.-M.; investigation, L.F.C.-M.; resources, L.F.C.-M.; data curation, Y.A.G.-G.; writing—original draft preparation, L.F.C.-M.; writing—review and editing, Y.A.G.-G.; visualization, L.F.C.-M.; supervision, Y.A.G.-G.; project administration, L.F.C.-M. 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:** The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

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

## References

1. Suárez, B.; Vera, J.D.; Botero, F.; Suárez, B.H.; Giraldo, W. Hidroituango Intake Gate Closure—Emergency Conditions. *Rev. Fac. Ing.* **2022**. [\[CrossRef\]](#)
2. Datta, A.; Maharaj, S.; Prabhu, G.N.; Bhowmik, D.; Marino, A.; Akbari, V.; Rupavatharam, S.; Sujeetha, J.A.R.P.; Anantrao, G.G.; Poduvattil, V.K.; et al. Monitoring the Spread of Water Hyacinth (*Pontederia crassipes*): Challenges and Future Developments. *Front. Ecol. Evol.* **2021**, *9*. [\[CrossRef\]](#)
3. Malik, A. Environmental Challenge Vis a Vis Opportunity: The Case of Water Hyacinth. *Environ. Int.* **2007**, *33*, 122–138. [\[CrossRef\]](#)
4. Barrett, S.C.H.; Forno, I.W. Style Morph Distribution in New World Populations of *Eichhornia crassipes* (Mart.) Solms-Laubach (*Water hyacinth*). *Aquat. Bot.* **1982**, *13*, 299–306. [\[CrossRef\]](#)
5. Wilson, J.R.; Holst, N.; Rees, M. Determinants and Patterns of Population Growth in Water Hyacinth. *Aquat. Bot.* **2005**, *81*, 51–67. [\[CrossRef\]](#)

6. Rodríguez-Lara, J.W.; Cervantes-Ortiz, F.; Arámbula-Villa, G.; Mariscal-Amaro, L.A.; Aguirre-Mancilla, C.L.; Andrio-Enríquez, E. Water Hyacinth (*Eichhornia crassipes*): A Review. *Agron. Mesoam.* **2022**, *33*. Available online: <https://www.redalyc.org/journal/437/43768481006/43768481006.pdf> (accessed on 1 September 2024).
7. Harun, I.; Pushiri, H.; Amirul-Aiman, A.J.; Zulkeflee, Z. Invasive Water Hyacinth: Ecology, Impacts and Prospects for the Rural Economy. *Plants* **2021**, *10*, 1613. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Pádua, L.; Antão-Geraldes, A.M.; Sousa, J.J.; Rodrigues, M.Â.; Oliveira, V.; Santos, D.; Miguens, M.F.P.; Castro, J.P. Water Hyacinth (*Eichhornia crassipes*) Detection Using Coarse and High Resolution Multispectral Data. *Drones* **2022**, *6*, 47. [\[CrossRef\]](#)
9. Thamaga, K.H.; Dube, T. Understanding Seasonal Dynamics of Invasive Water Hyacinth (*Eichhornia crassipes*) in the Greater Letaba River System Using Sentinel-2 Satellite Data. *GLSCI Remote Sens.* **2019**, *56*, 1355–1377. [\[CrossRef\]](#)
10. Godana, G.; Fufa, F.; Debesa, G. Eichhornia Crassipes Expansion Detection Using Geospatial Techniques: Lake Dambal, Oromia, Ethiopia. *Environ. Chall.* **2022**, *9*, 100616. [\[CrossRef\]](#)
11. Villa, P.; Bresciani, M.; Bolpagni, R.; Pinardi, M.; Giardino, C. A Rule-Based Approach for Mapping Macrophyte Communities Using Multi-Temporal Aquatic Vegetation Indices. *Remote Sens. Environ.* **2015**, *171*, 218–233. [\[CrossRef\]](#)
12. Pádua, L.; Duarte, L.; Antão-Geraldes, A.M.; Sousa, J.J.; Castro, J.P. Spatio-Temporal Water Hyacinth Monitoring in the Lower Mondego (Portugal) Using Remote Sensing Data. *Plants* **2022**, *11*, 3465. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Belgiu, M.; Drăguț, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [\[CrossRef\]](#)
14. Mukarugwiro, J.A.; Newete, S.W.; Adam, E.; Nsanganwimana, F.; Abutaleb, K.; Byrne, M.J. Mapping Spatio-Temporal Variations in Water Hyacinth (*Eichhornia crassipes*) Coverage on Rwandan Water Bodies Using Multispectral Imageries. *Int. J. Environ. Sci. Technol.* **2021**, *18*, 275–286. [\[CrossRef\]](#)
15. BID. *EPM Implementacion Modelo Calidad del Agua Phi Resumen Didáctico*; BID: Medellín, Colombia, 2016.
16. Roy, D.P.; Li, J.; Zhang, H.K.; Yan, L. Best Practices for the Reprojection and Resampling of Sentinel-2 Multi Spectral Instrument Level 1C Data. *Remote Sens. Lett.* **2016**, *7*, 1023–1032. [\[CrossRef\]](#)
17. Rolon-Mérettea, D.; Ross, M.; Rolon-Mérettea, T.; Church, K. Introduction to Anaconda and Python: Installation and setup. *Python Res. Psychol.* **2020**, *16*, S3–S11. [\[CrossRef\]](#)
18. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
19. Liaw, A.; Wiener, M. Classification and Regression by Random Forest. *R News* **2002**, *2*, 18–22.
20. Gislason, P.O.; Benediktsson, J.A.; Sveinsson, J.R. Random Forests for Land Cover Classification. *Pattern Recognit. Lett.* **2006**, *27*, 294–300. [\[CrossRef\]](#)
21. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [\[CrossRef\]](#)
22. Bayable, G.; Cai, J.; Mekonnen, M.; Legesse, S.A.; Ishikawa, K.; Imamura, H.; Kuwahara, V.S. Detection of Water Hyacinth (*Eichhornia crassipes*) in Lake Tana, Ethiopia, Using Machine Learning Algorithms. *Water* **2023**, *15*, 880. [\[CrossRef\]](#)
23. Powers, D.M.W. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation. *arXiv* **2020**, arXiv:2010.16061. [\[CrossRef\]](#)
24. Dersseh, M.G.; Steenhuis, T.S.; Kibret, A.A.; Eneyew, B.M.; Kebedew, M.G.; Zimale, F.A.; Worqlul, A.W.; Moges, M.A.; Abebe, W.B.; Mhret, D.A.; et al. Water Quality Characteristics of a Water Hyacinth Infested Tropical Highland Lake: Lake Tana, Ethiopia. *Front. Water* **2022**, *4*, 774710. [\[CrossRef\]](#)
25. Avci, C.; Budak, M.; Yağmur, N.; Balçık, F. Comparison between Random Forest and Support Vector Machine Algorithms for LULC Classification. *Int. J. Eng. Geosci.* **2023**, *8*, 1–10. [\[CrossRef\]](#)

**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.