ABES

Revista Brasileira de Ciências Ambientais
Brazilian Journal of Environmental Sciences

Detection of *Eichhornia crassipes* and *Pistia stratiotes* in the São Francisco River Basin via Remote Sensing

Detecção de *Eichhornia crassipes* e *Pistia stratiotes* na Bacia do Rio São Francisco via Sensoriamento Remoto

Luiz Felipe Naziazeno Neto[1], *Luiz Filipe Barbosa Varjão*[1], *Jarbas José de Souza Silva*[1], *Yan Felipe Pereira Fernandes*[1], *Évelyn Marcia Pôssa*[1], *Ronny Francisco Marques de Souza*[1]

ABSTRACT

Infestations of *Eichhornia crassipes* (Mart.) Solms and *Pistia stratiotes* L. compromise the ecological equilibrium and the safety of water management within the São Francisco River Basin. This study evaluated the performance of Random Forest (RF), Support Vector Machine (SVM), and Classification and Regression Trees (CART) algorithms for mapping these invasive macrophytes in the sub-middle region of the basin (Jatobá, Pernambuco, Brazil). Five land cover classes were identified using Landsat 8 imagery and Normalized Difference Vegetation Index (NDVI) datasets processed in QGIS 3.16. RF and CART displayed identical overall accuracy rates of 94%, and all three algorithms achieved 100% accuracy in isolating *Pistia stratiotes.* Conversely, persistent spectral confusion between *Eichhornia crassipes* and soil led the RF model to misclassify 33% of soil points due to spatial resolution limitations (30 m). SVM achieved an overall accuracy of 88%, yielding a high recall rate (86%) for water hyacinth but overestimating its area. The results validate orbital remote sensing coupled with machine learning as an effective architecture for continuous bioinvasion monitoring in target zones of the Brazilian semiarid region.

Keywords: bioinvasion; Landsat 8; machine learning; predictive modeling.

RESUMO

Infestações de *Eichhornia crassipes* (Mart.) Solms e *Pistia stratiotes* L. comprometem o equilíbrio ecológico e a segurança da gestão hídrica na Bacia do Rio São Francisco. Este estudo avaliou o desempenho dos algoritmos Random Forest (RF), Support Vector Machine (SVM) e Classification and Regression Trees (CART) no mapeamento dessas macrófitas invasoras na região do submédio da bacia (Jatobá, Pernambuco-Brasil). Cinco classes de cobertura da terra foram identificadas a partir de imagens Landsat 8 e dados de Índice de Vegetação Normalizada processados no QGIS 3.16. O RF e o CART apresentaram índices idênticos de acurácia global (94%), e todos os três algoritmos obtiveram 100% de acerto no isolamento da *Pistia stratiotes*. Por outro lado, ocorreu uma confusão espectral persistente entre a *Eichhornia crassipes* e o solo, levando o modelo RF a classificar incorretamente 33% dos pontos de solo devido às limitações de resolução espacial (30 m). O SVM alcançou acurácia global de 88%, garantindo alta taxa de sensibilidade (86%) para o aguapé, embora tenha superestimado sua área. Os resultados validam o sensoriamento remoto orbital aliado ao aprendizado de máquina como uma arquitetura eficaz para o monitoramento contínuo da bioinvasão em áreas estratégicas do semiárido brasileiro.

Palavras-chave: bioinvasão; Landsat 8; aprendizado de máquina; modelagem preditiva.

[1]Instituto Federal de Alagoas – Piranhas (AL), Brazil.

Corresponding author: Ronny Francisco Marques de Souza – Instituto Federal de Alagoas (IFAL), Campus Piranhas – Av. Sergipe, s/n – Piranhas (AL), CEP 57460-000, Brazil. E-mail: ronny.souza@ifal.edu.br

Received on: 06/21/2025. Accepted on: 09/04/2026.

https://doi.org/10.67790/rbciamb.026.2662

Introduction

The arrival of exotic species in new environments has become a serious problem. Their relationship with the disappearance of native organisms, a process known as extinction, is now widely accepted and documented (Haubrock et al., 2024). The dispersal and extinction of species are natural processes, but their rates have been profoundly accelerated by humans throughout history (Riahi, 2020). The establishment of the first human groups enabled the exchange of materials and species, creating incipient exchange networks that expanded dramatically with the onset of early global trade, a historical phenomenon conceptualized as the Columbian Exchange (Crosby, 2004). Once these exotic organisms became successfully established in new territories, the problems related to invasive species began to attract attention (Amorim et al., 2024). When they exhibit high ecological plasticity, such organisms become aggressive biological invaders (Haubrock et al., 2025). Within freshwater ecosystems, two invasive aquatic macrophytes from South America exhibit exceptional propagation capabilities and stand out as major ecological threats: *Eichhornia crassipes* (Mart.) Solms (water hyacinth) and *Pistia stratiotes* L. (water lettuce) (Macêdo et al., 2024).

Eichhornia crassipes and *Pistia stratiotes* are free-living aquatic macrophytes native to South American basins and belonging to the Pontederiaceae and Araceae families, respectively. Both species reproduce sexually and asexually, using clonal strategies for rapid propagation and showing high tolerance to wide variations in abiotic conditions, such as temperature, pH, and salinity. Consequently, they form dense, floating green mats on the water, causing unprecedented environmental, economic, and social problems (Villamagna and Murphy, 2010). Their dispersion in continental waters worldwide, initially associated with landscaping, is accelerated in anthropogenically eutrophicated aquatic environments with high concentrations of phosphorus and nitrogen (Petrúcio and Esteves, 2000), especially in tropical and subtropical regions (CABI, 2022). Currently, these macrophytes pose severe biological risks, and the multifaceted ecological damage driven by their aggressive exponential growth patterns has been extensively documented (Macêdo et al., 2024). Although these invaders do not act as primary drivers of eutrophication, their continuous growth and decomposition cycles significantly increase the internal organic matter load, thereby accelerating oxygen depletion and further degrading aquatic ecosystems (Villamagna and Murphy, 2010). Conversely, under controlled conditions, their high nutrient uptake capacity enables them to be used as effective agents for environmental phytoremediation (Alexandre et al., 2025).

In Brazil, the dispersion of these exotic species beyond their natural geographic limits has reached all major river basins, including the São Francisco River (Oliveira et al., 2005). The São Francisco River Basin is one of the most important systems in the country, covering approximately 500 municipalities across seven states (ANA, 2022). It holds immense economic and strategic relevance owing to hydroelectric energy production, public water supply, and irrigation in the semiarid Northeast, in addition to hosting important gastronomic and tourist destinations (CBHSF, 2016). To monitor these vast areas, remote sensing technologies enable a spatiotemporal understanding where traditional visual methods fail (Thamaga and Dube, 2018). The application of these tools has become highly viable due to Earth observation programs, such as the Sentinel missions (European Space Agency) and the Landsat Program (NASA/USGS), which have free data distribution policies, enabling regular biomass mapping without prohibitive costs (Souza et al., 2024).

The study area is the municipality of Jatobá (PE), Brazil, a strategically vital region for national energy production, ensuring electricity supply and development for the Northeast. Located between the Luiz Gonzaga and Paulo Afonso hydroelectric power plants, the area has a combined installed capacity of approximately 5.7 GW (CHESF, 2022). Furthermore, intensive local aquaculture positions Jatobá as the largest tilapia (*Oreochromis niloticus*) producer in the Northeast and the third-largest in the country (IBGE, 2023). Local environmental features, such as high nutrient loads from human activities (fish farms, urban waste, agricultural runoff) and the region's lentic flow system, make the region conducive to the development and spread of these invasive macrophytes (Carvalho et al., 2023).

In recent years, the use of machine learning methods to track aquatic weeds has expanded through diverse computational strategies. Historically, Dube et al. (2017) reported excellent baseline results, achieving a 95% classification rate for water hyacinth using Landsat 8 data processed with discriminant analysis frameworks. Since then, evaluating which algorithm best adapts to different regional contexts has become standard practice in the literature. For instance, Mukarugwiro et al. (2019) tested random forest (RF) and support vector machine (SVM) models on Landsat 8 images over water bodies in Rwanda, observing that the decision-tree approach achieved 85% accuracy, significantly outperforming vector-space classification, which reached 65%. This general trend was later reinforced by Pádua et al. (2022), who used orbital and drone imagery and found that RF achieved 94% accuracy, outperforming SVM. Working along similar methodological lines, Ade et al. (2022) achieved 90% overall accuracy with Sentinel-2 datasets, while Singh et al. (2020) demonstrated the strength of hierarchical classification by achieving a 93% success rate. More recently, a broad seasonal study by Bayable et al. (2023) using the Google Earth Engine platform confirmed that both Landsat and Sentinel sensors allow these predictive models to surpass 90% accuracy at different times of the year, with RF consistently delivering the most balanced statistical indicators. Current efforts in the field are now moving toward integrating optical data with radar sensors to better assess how these plants develop over time (Belayhun et al., 2024; Kinara et al., 2024). However, a common limitation across this body of work is a heavy reliance on isolating a single target species, usually water hyacinth. This creates a clear gap in semiarid regions (Almeida et al., 2023), where different invasive plants often coexist and require a system capable of mapping multiple classes simultaneously.

In this context, this research evaluated an aquatic plant detection protocol using Landsat 8 satellite imagery and machine learning architectures. The core Pynton algorithm used for image processing and classification was registered with the National Institute of Industrial Property (INPI) under registration number BR512024002605-3 to secure computational copyright. However, semiarid water bodies frequently present co-occurring invasive populations that require a simultaneous multi-class approach. While the INPI registration protects the software code, the primary scientific contribution of this setup is the integration of datasets calibrated to address localized spectral overlaps among multiple coexisting macrophytes in a semiarid lentic system. Therefore, the objective of this study was to compare the performance of RF, SVM, and Classification and Regression Trees (CART) algorithms for the simultaneous detection of *Eichhornia crassipes* and *Pistia stratiotes* within the sub-middle region of the São Francisco River Basin.

Methodological procedures

The computational and spatial framework proposed in this study establishes a machine-learning routine for the simultaneous detection and mapping of *Eichhornia crassipes* and *Pistia stratiotes* within a highly dynamic semiarid aquatic ecosystem. The methodological architecture integrates digital image processing, spectral index extraction (the Normalized Difference Vegetation Index, NDVI), and supervised classification techniques to evaluate the predictive performance of tree-based and vector-space algorithms. Rather than treating the analysis as a single-species problem, the protocol is structured as a robust multi-class assessment calibrated to handle localized spectral overlaps, thereby enhancing natural resource management.

Data acquisition and study area

Geographical characterization of the study area

The study area is situated in the Brazilian municipality of Jatobá, located in the state of Pernambuco. This region is characterized by gently undulating terrain and features hyper-xerophytic vegetation, indicative of its semiarid climate. The precise geographical delimitation of this study area is presented in Figure 1.

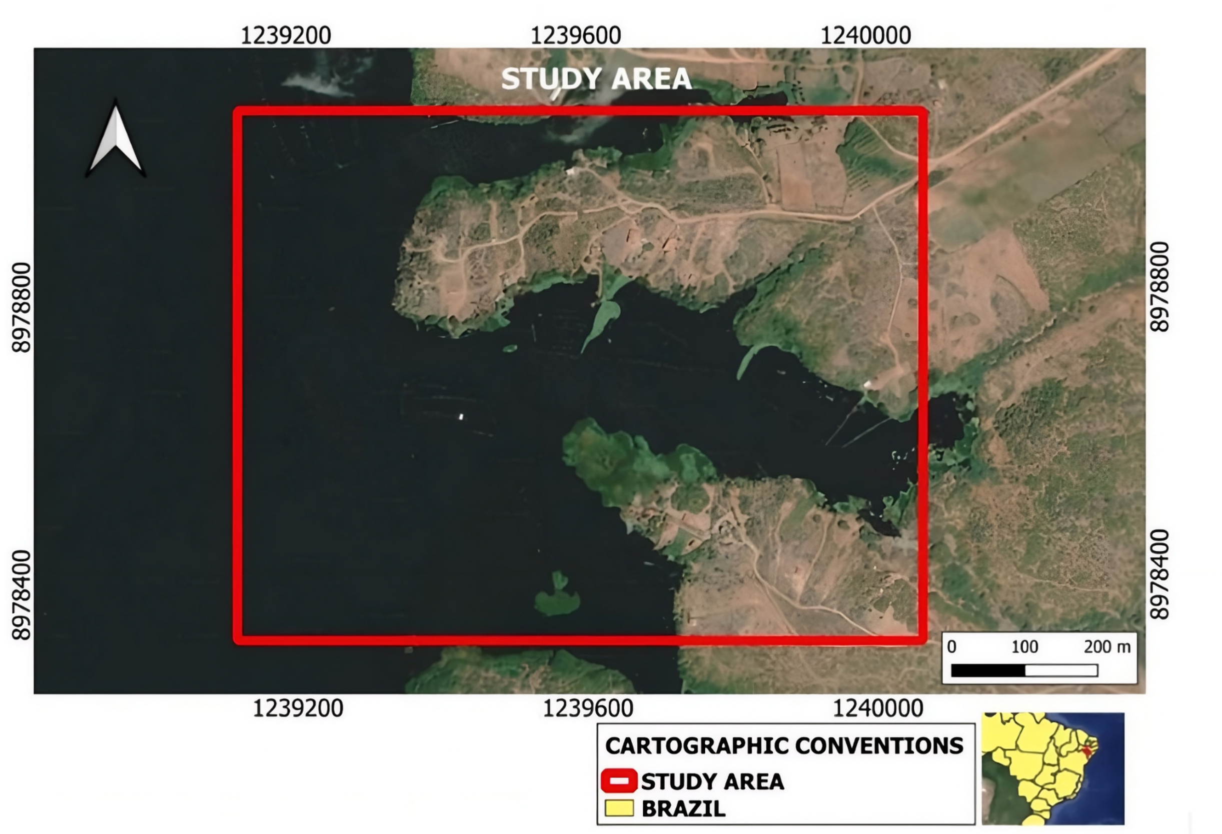

Figure 1 – Study area map
Source: Google Satellite Database.

Satellite data acquisition and characteristics

Satellite images were obtained from the United States Geological Survey (USGS) platform (USGS, 2022) using data from the Landsat 8 satellite, operated by NASA. This Earth observation platform provides medium-resolution multispectral data. Landsat 8 images feature a 30-meter spatial resolution, which is highly compatible with analyses of regional land cover and vegetation dynamics. The ability of Landsat 8 to capture data across multiple spectral bands, including the Near-Infrared (NIR) band, which is indispensable for calculating the NDVI, combined with its extensive temporal data series, makes it a robust tool for long-term environmental monitoring and invasive species detection without prohibitive operational costs. To facilitate systematic processing, the raw geospatial data structures were managed using cloud-based data repositories.

Sample collection and categorization

For model training and validation, a field visit to the study area was conducted, complemented by a detailed visual inspection of satellite images. This hybrid process enabled the collection of 500 georeferenced ground-truth points. These points were categorized into five main land cover classes: Water, Exposed Soil, *Eichhornia crassipes*, *Pistia stratiotes*, and Native Forest. Organizing these data into shapefiles in QGIS 3.16 facilitated the identification and extraction of the target radiometric responses. The importance of these samples lies in their ability to accurately represent different land cover types, which is essential for the quality and precision of machine learning predictions (Yang et al., 2020). Representative samples of each of these five classes are presented in Figure 2.

Detection model development

Image processing and normalization

Digital image processing and spectral normalization were executed through an integrated computational workflow. The QGIS 3.16 software was used solely for geospatial data manipulation, vector boundary definition, and visual inspection of surface samples. Subsequently, the pre-processed Landsat 8 bands were integrated into a cloud-based Python environment to run the proprietary automated routine registered with the INPI. Within this programmatic framework, spectral matrix normalization and the mathematical extraction of the NDVI were performed directly in code, leveraging the Near-Infrared (B5) and Red (B4) spectral channels to optimize data ingestion for machine learning classifiers (Bayable et al., 2023; NASA, 2025). To eliminate radiometric inconsistencies and stabilize algorithmic training, a linear min-max normalization protocol was systematically applied, rescaling all raw index features to a strict operational range of -1 to 1 to ensure data uniformity across the multitemporal series (Song et al., 2020).

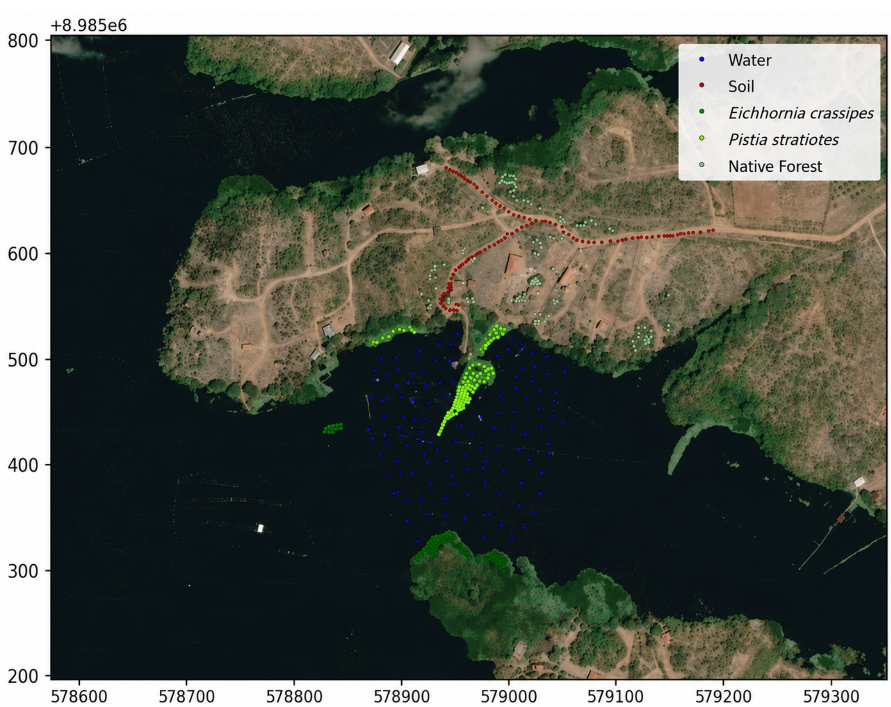

Figure 2 – Class samples

Algorithmic training and classification framework

The predictive evaluation of the tree-based and vector-space architectures was performed by ingesting the normalized NDVI matrices and class labels into the automated Python routine. The training phase systematically compared three supervised predictive models: RF, SVM, and CART. The input data matrix was partitioned using a cross-validation protocol to stabilize hyperparameter optimization and prevent model overfitting.

For the RF implementation, the ensemble structure was computed by aggregating multiple randomized decision-tree estimators to minimize global variance (Breiman, 2001). The SVM setup was configured using a radial basis function (RBF) kernel matrix to resolve localized spectral nonlinearities and nonseparable class boundaries between vegetation types (Cortes and Vapnik, 1995). Lastly, the CART architecture was implemented as a nonparametric single-tree baseline, establishing hierarchical decision rules based on feature thresholds derived from the training data (Quinlan, 1986). Computational execution and the extraction of performance metrics for all three algorithms were managed with the scikit-learn library framework (Pedregosa et al., 2011).

Predictive performance assessment

Following the training phase, the validation matrices were subjected to statistical analysis to assess the operational effectiveness of the RF, SVM, and CART models. The spatial application of the trained algorithms over the multitemporal raster scenes enabled pixel-by-pixel classification. Model performance was quantitatively evaluated using a comprehensive suite of statistical metrics, including overall accuracy, precision, and recall. These metrics were computed on the validation sub-dataset to measure the proportion of correct global classifications, the rate of true positive assignments relative to false positives (commission errors), and the models' ability to capture target instances relative to false negatives (omission errors).

To detect localized classification overlaps and systematic displacements, confusion matrices were systematically generated. These matrices provided a diagnostic visualization of cross-class performance, isolating specific error patterns and detailing each algorithm's capacity to distinguish between conflicting surface classes, such as native vegetation, exposed soil, and the target aquatic invasive species (Ma et al., 2020).

Results and discussion

The spatial classification models generated through the computational routine yielded specific metrics for tracking the co-occurrence of *Eichhornia crassipes* and *Pistia stratiotes* within the study area. In terms of overall accuracy, both the ensemble RF and the hierarchical CART architectures achieved 94%, outperforming the SVM baseline, which reached 88%. This divergence suggests that tree-based nonparametric structures better adapt to the asymmetric spectral distributions typically observed in semiarid regions.

Regarding statistical precision, RF achieved 94.32%, followed by CART at 94.23% and SVM at 91.95%. These precision values reflect the tree-based model's ability to mitigate commission errors, reducing false positives in critical areas. In practical terms, avoiding the misclassification of clean water bodies or native Caatinga margins as invasive biomass is relevant for environmental enforcement and local economic activities, preventing operational disruptions in Jatobá's aquaculture sector and in the hydroelectric reservoirs.

For recall, all three algorithms achieved 94%. This parity indicates that, despite variations in overall accuracy, all models maintained the sensitivity required to detect the presence of biological invaders, minimizing omission errors for early-warning ecological monitoring. The observed predictive behavior aligns with previous remote sensing literature, which notes the performance of tree-based methods when processing multispectral scenes in complex environments (Mukarugwiro et al., 2019; Pádua et al., 2022; Bayable et al., 2023). The spatial distribution of these thematic classifications across the geographic subset is illustrated in Figure 3.

The performance of the RF and CART models can be attributed to the ability of decision-tree structures to resolve nonlinear spectral relationships and handle multi-class imbalances without requiring rigid parametric distributions. The hierarchical splitting mechanism of CART architectures generates explicit rule-based thresholds for environmental targets. Furthermore, the ensemble randomization mechanism inherent to RF execution reduces structural variance and mitigates overfitting, stabilizing predictions across heterogeneous landscapes (Breiman, 2001). Conversely, while the SVM algorithm yielded inferior global metrics compared to the tree-based frameworks, its vector-space separation protocol remained stable, using kernel functions to define optimal separating hyperplanes between overlapping radiometric profiles (Cortes and Vapnik, 1995).

To verify the localized classification performance under independent operational conditions, the confusion matrix for the RF model was computed from the validation dataset (Table 1). The cross-tabulation analysis demonstrated the model's ability to resolve spectral classes across the study area.

The cross-tabulation matrix details the internal balance of the ensemble model, highlighting its discriminant characteristics and localized radiometric challenges (Table 1). The computational routine achieved a classification accuracy of 100% for *Pistia stratiotes*, along with separation performance of 99% for Native Forest and 94% for Water. These values indicate that the algorithmic structure extracted distinct phenological and moisture-related signatures characterizing these surfaces, thereby reducing spatial overestimation in open-water regions and dense terrestrial canopies.

Conversely, a spectral overlap occurred between the Exposed Soil and *Eichhornia crassipes* domains. The model exhibited a 33% com-

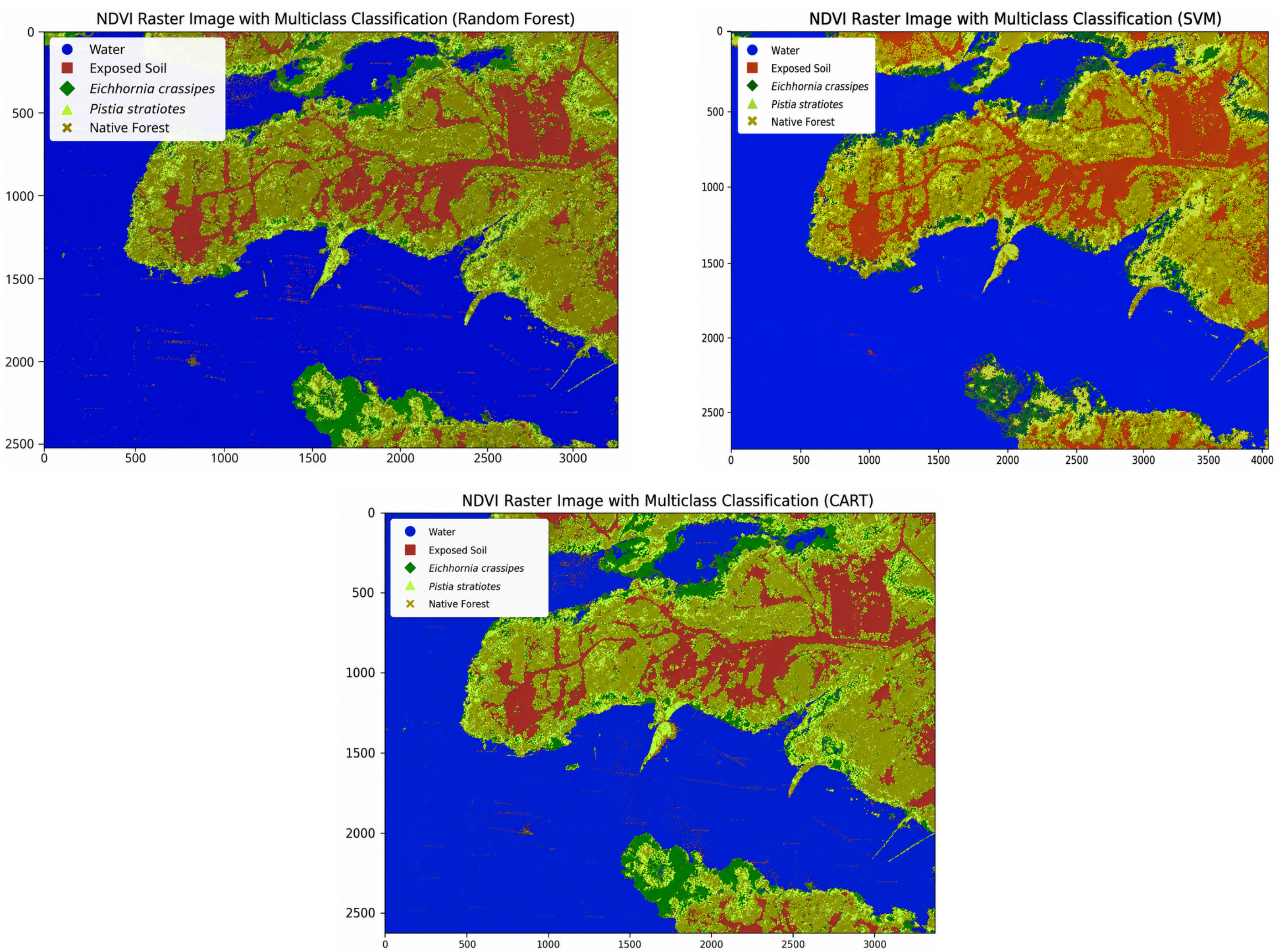

Figure 3 – Images generated with the RF, SVM, and CART algorithms

Table 1 – Confusion matrix (%) for RF

Class	Water	Exposed Soil	*E. crassipes*	*P. stratiotes*	Native Forest
Water	94	2	1	0	3
Exposed Soil	2	65	33	0	0
E. crassipes	0	17	83	0	0
P. stratiotes	0	0	0	100	0
Native Forest	1	0	0	0	99

mission error, with exposed soil pixels misclassified as water hyacinth, and a 17% omission error, shifting aquatic biomass toward the soil class. Within the semiarid geographic context of the sub-middle São Francisco River Basin, this behavior is associated with the 30-meter spatial resolution of the sensor. During seasonal low-water periods, marginal macrophyte banks undergo partial desiccation, accumulate organic detritus, and transition into sub-pixel mixtures in which the bare-soil background and dry biomass merge into an overlapping radiometric profile. Additionally, suspended sediment loads in shallow littoral zones can mimic the reflectance responses of mixed edge pixels. These localized border confusions indicate the potential value of integrating multi-sensor frameworks into future high-resolution monitoring architectures.

To evaluate the variables that contributed to the overall accuracy of 88% observed with the vector-space classifier, the SVM's behavior was tabulated (Table 2).

The matrix shows that while the SVM isolated the biological invaders, achieving 100% classification accuracy for *Pistia stratiotes* and 86% for *Eichhornia crassipes*, variations occurred when processing the background control classes.

An operational variation is evidenced by the misclassification of 59% of open-water pixels as Native Forest, alongside a 51% commission displacement that shifted Exposed Soil into the *Eichhornia crassipes* domain. Within the sub-middle São Francisco River environment, this confusion indicates that the radial basis function kernel boundaries overfitted to the NIR and localized NDVI responses of dense vegetation. Consequently, the algorithm did not establish an operational margin for shallow, sediment-rich littoral waters and marginal dry soils, and it projected their mixed radiometric profiles into terrestrial and aquatic biomass categories. This performance reflects the limitations imposed by vector-space hyperplanes when handling multitemporal scenes under semiarid conditions, indicating that single-tree and ensemble architectures provide alternative frameworks for regional environmental management.

To evaluate the behavior of a nonparametric single-tree architecture, the classification performance of the CART model was tabulated (Table 3).

The CART model achieved 98% accuracy in isolating open-water bodies and maintained 100% classification accuracy when detecting *Pistia stratiotes*. This indicates that for differentiated radiometric targets, a single nonparametric tree can compute threshold splits that equal the performance of ensemble structures.

During the transitional dynamics of semiarid riparian margins, variations in the single-tree architecture occurred. The model replicated the spectral overlap observed in the alternative tree-based framework, showing a 31% commission error, with Exposed Soil misclassified as *Eichhornia crassipes*, and a 19% omission error, with water hyacinth misclassified as soil. Because this confusion pattern persists across both RF and CART architectures, it indicates that the mixed-pixel phenomenon at a 30-meter spatial resolution, driven by marginal desiccation and fluctuating water levels, is associated with environmental and sensor characteristics.

A comparative inspection reveals that CART detected *Eichhornia crassipes* at 80%, whereas the RF model achieved 83%. This 3% difference in the main target invasive class indicates the greater sensitivity of a single tree to localized radiometric noise. By not using randomized feature subsets, CART establishes orthogonal boundaries that make sensitivity less stable under varying vegetation conditions. This variation indicates the greater stability of the RF ensemble in the automated computational routine, providing a framework for multi-class classification precision.

Study limitations and future perspectives

The operational parameters of this study reflect the limitations imposed by the 30-meter spatial resolution of the Landsat 8 sensor, which influences the detection of marginal macrophyte banks during transitional seasonal periods. Additionally, the modeling framework used the NDVI metric as the primary vegetation index, representing an approach based on spectral boundaries. The performance metrics obtained characterize the local geographic variables of the sub-middle São Francisco River Basin. Future applications can evaluate the integration of multi-sensor architectures, incorporating platforms such as Sentinel-2, to analyze multi-class dynamics at other scales.

The comparative evaluation of RF, SVM, and CART models demonstrated the applicability of machine learning to *Eichhornia*

Table 2 – Confusion matrix (%) for SVM

Class	Water	Exposed Soil	*E. crassipes*	*P. stratiotes*	Native Forest
Water	37	0	4	0	59
Exposed Soil	7	42	51	0	0
E. crassipes	0	14	86	0	0
P. stratiotes	0	0	0	100	0
Native Forest	1	0	0	0	99

Table 3 – Confusion matrix (%) for CART

Class	Water	Exposed Soil	*E. crassipes*	*P. stratiotes*	Native Forest
Water	98	1	0	0	1
Exposed Soil	3	66	31	0	0
E. crassipes	1	19	80	0	0
P. stratiotes	0	0	0	100	0
Native Forest	3	0	0	0	97

crassipes detection. All three models classified spectrally distinct classes, such as *Pistia stratiotes* and Native Forest, without error. RF and CART achieved an overall accuracy of 94%. CART detected *E. crassipes* with fewer false positives from other classes, while RF offered an operational balance. Although SVM achieved an accuracy of 88%, it had a recall of 86% for *E. crassipes* detection, indicating that it correctly identified the invasive species' presence. The confusion with Exposed Soil occurred across the tested architectures. These models provide data for management and conservation strategies.

The comparative evaluation of the model developed for *Eichhornia crassipes* detection indicates characteristics comparable to those reported in previous studies. With RF and CART achieving 94% overall accuracy, the results are close to those reported by Dube et al. (2017), who obtained 95% using Landsat 8, and Bayable et al. (2023), who reported over 90% accuracy with the same sensor, while differing from the metrics presented by Mukarugwiro et al. (2019) and Ade et al. (2022). The performance of RF and CART relative to SVM (88% accuracy) is consistent with observations reported by Pádua et al. (2022) and Mukarugwiro et al. (2019).

A characteristic of the proposed framework is 100% classification accuracy for *Pistia stratiotes* across the algorithms. For *E. crassipes*, although confusion with Exposed Soil occurred, the models displayed distinct operational parameters: SVM achieved 86% recall for this class, followed by RF (83%) and CART (80%), with variations in false-positive rates. This classification behavior for both invasive species in the São Francisco River relates to its practical application for environmental monitoring and decision-making. The applicability of this approach was evaluated in the sub-middle region of the São Francisco River, an area characterized by multiple activities, including energy production, fish farming, water supply, navigation, and tourism, which are affected by the proliferation of both species, providing data to support the mitigation of these regional environmental and socioeconomic impacts.

Conclusion

The models evaluated in this study, using Landsat 8 imagery and machine learning algorithms integrated with the NDVI metric, detected the presence of *Eichhornia crassipes* and *Pistia stratiotes*. The data obtained indicate the distribution of these macrophytes, supporting the selection of targeted monitoring parameters. This architecture demonstrated a framework for dual invasive species detection, classifying both invaders across the monitored sectors of the sub-middle São Francisco River region.

In summary, this application indicates the potential of computational frameworks in environmental remote sensing, demonstrating that automated classification routines can serve as tools to support data generation for regional water resource management and conservation policies.

Authors' Contributions: Naziazeno Neto, L.F.: Software, Methodology, Formal Analysis, Writing, original draft; Varjão, L.F.B.: Data curation, Investigation; Silva, J.J.S.: Data curation, Investigation; Fernandes, Y.F.P.: Data curation, Investigation; Pôssa, E.M.: Methodology, Validation, Visualization, Writing, review & editing; Souza, R. F. M.: Conceptualization, Supervision, Project administration, Resources, Writing, review & editing.

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

Funding: National Council for Scientific and Technological Development (CNPq).

Declaration on the use of generative artificial intelligence: During the preparation of this manuscript, the authors used Gemini (Google) for polishing the scientific English, improving conciseness, and ensuring conceptual consistency during the peer-review revision stage. After using this tool, the authors reviewed and edited the content as necessary and take full responsibility for the content of the published article.

References

Ade, C.; Khanna, S.; Lay, M.; Ustin, S.L.; Hestir, E.L., 2022. Genus-level mapping of invasive floating aquatic vegetation using Sentinel-2 satellite remote sensing. Remote Sensing, v. 14 (13), 3013. https://doi.org/10.3390/rs14133013

Agência Nacional de Águas e Saneamento Básico (ANA), 2022. Relatório de Conjuntura dos Recursos Hídricos no Brasil 2022. ANA, Brasília.

Alexandre, D.S.; Ogura, A.P.; Mohedano, R.A.; Thibau, L.B.G., 2025. Potential of two floating aquatic macrophytes in improving water quality: a case study in two tropical streams. Revista Brasileira de Ciências Ambientais (RBCIAMB), v. 60, e2158. https://doi.org/10.5327/Z2176-94782158

Almeida, D.N.O.; Araújo, D.C.S.; Soares, D.R.; Maia, F.M.A.; Montenegro, S.M.G.L. et al., 2023. Delimitation of water areas using remote sensing in Brazil's semiarid region. Revista Brasileira de Ciências Ambientais (RBCIAMB), v. 58 (1), 20-29. https://doi.org/10.5327/Z2176-94781524

Amorim, N.D.B.O.; Oliveira, I.; Liesenfeld, M.V.A., 2024. Popular knowledge and perceptions of invasive exotic species in Acre, Brazil. Revista Brasileira de Ciências Ambientais (RBCIAMB), v. 59, e1762. https://doi.org/10.5327/Z2176-94781762

Bayable, G.; Cai, J.; Mekonnen, M.; Legesse, S.A.; Ishikawa, K. et al., 2023. Detection of water hyacinth (*Eichhornia crassipes*) in Lake Tana, Ethiopia, using machine learning algorithms. Water, v. 15 (5), 880. https://doi.org/10.3390/w15050880

Belayhun, M.; Chere, Z.; Abay, N.G.; Nicola, Y.; Asmamaw, A., 2024. Spatiotemporal pattern of water hyacinth (*Pontederia crassipes*) distribution in Lake Tana, Ethiopia, using a random forest machine learning model. Frontiers in Environmental Science, v. 12, 1476014. https://doi.org/10.3389/fenvs.2024.1476014

Breiman, L., 2001. Random forests. Machine Learning, v. 45 (1), 5-32. https://doi.org/10.1023/A:1010933404324

CABI Compendium International, 2022. *Eichhornia crassipes* (water hyacinth). https://doi.org/10.1079/cabicompendium.20544

Carvalho, B.C.; Colares, M.F.A.; Barbosa, C.C.; Figueira, R.M.A.; Fortunato, H.F.M., 2023. Challenges and advances of inland waters monitoring: a case study highlighting the use of satellite imagery and cloud computing in Brazil. Limnology and Oceanography Bulletin, v. 32 (4), 129-135. https://doi.org/10.1002/lob.10600

Comitê da Bacia Hidrográfica do Rio São Francisco (CBHSF), 2016. Plano de Recursos Hídricos da Bacia Hidrográfica do Rio São Francisco (2016-2025): resumo executivo. CBHSF, Belo Horizonte.

Companhia Hidro Elétrica do São Francisco (CHESF), 2022. Relatório anual 2022. CHESF, Recife.

Cortes, C.; Vapnik, V., 1995. Support-vector networks. Machine Learning, v. 20 (3), 273-297. https://doi.org/10.1007/BF00994018

Crosby, A.W., 2004. Ecological Imperialism: The Biological Expansion of Europe, 900-1900. 2nd ed. Cambridge University Press, Cambridge.

Dube, T.; Mutanga, O.; Sibanda, M.; Bangamwabo, V.; Shoko, C., 2017. Testing the detection and discrimination potential of the new Landsat 8 satellite data on the challenging water hyacinth (*Eichhornia crassipes*) in freshwater ecosystems. Applied Geography, v. 84, 11-22. https://doi.org/10.1016/j.apgeog.2017.04.005

Haubrock, P.J.; Soto, I.; Ahmed, D.A.; Ansri, R.A.; Tarkan, A.S. et al., 2024. Biological invasions are a population-level rather than a species-level phenomenon. Global Change Biology, v. 30, e17312. https://doi.org/10.1111/gcb.17312

Haubrock, P.J.; Everts, T.; Abreo, N.A.S.; Bojko, J.; Deklerck, V. et al., 2025. The impacts of biological invasions. Biological Reviews. https://doi.org/10.1002/brv.70124

Instituto Brasileiro de Geografia e Estatística (IBGE), 2023. Produção da Pecuária Municipal 2023. IBGE, Rio de Janeiro.

Kinara, M.; Hadullo, K.; Mgala, M., 2024. Machine learning model for mapping and monitoring of water hyacinth in Lake Victoria, Kenya: a literature review. IOSR Journal of Engineering, v. 13 (8), 103-108. https://doi.org/10.9790/1813-1308103108

Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yi, G. et al., 2020. Deep learning in remote sensing applications: a meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing, v. 8 (3), 154-172. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Macêdo, R.L.; Haubrock, P.J.; Klippel, G.; Fernandez, R.D.; Leroy, B. et al., 2024. The economic cost of invasive aquatic plants: a global perspective on ecology and management gaps. Science of the Total Environment, v. 908 (15), 168217. https://doi.org/10.1016/j.scitotenv.2023.168217

Mukarugwiro, J.; Newete, S.; Adam, E.; Nsanganwimana, F.; Abutaleb, K. et al., 2019. Mapping distribution of water hyacinth (*Eichhornia crassipes*) in Rwanda using multispectral remote sensing imagery. African Journal of Aquatic Science, v. 44, 339-348. https://doi.org/10.2989/16085914.2019.1650712

National Aeronautics and Space Administration (NASA), 2025. NDVI - Normalized Difference Vegetation Index. https://earthobservatory.nasa.gov/features/MeasuringVegetation/measuring_vegetation_2.php

Oliveira, N.M.B.; Sampaio, E.V.S.B.; Pereira, S.M.B.; Junior, A.M.M., 2005. Regeneration capacity of *Egeria densa* in reservoirs in Paulo Afonso, Bahia. Planta Daninha, v. 23, 363-369. https://doi.org/10.1590/S0100-83582005000200026

Pádua, L.; Antão-Geraldes, A.M.; Sousa, J.J.; Rodrigues, M.; Oliveira, V. et al., 2022. Water hyacinth (*Eichhornia crassipes*) detection using coarse and high-resolution multispectral data. Drones, v. 6 (2), 47. https://doi.org/10.3390/drones6020047

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Thirion, B.; Grisel, O. et al., 2011. Scikit-learn: machine learning in Python. Journal of Machine Learning Research, v. 12, 2825-2830.

Petrúcio, M.M.; Esteves, F.A., 2000. Uptake rates of nitrogen and phosphorus by *Eichhornia crassipes* and *Salvinia auriculata* in the laboratory. Brazilian Journal of Biology, v. 60 (2), 229-236. https://doi.org/10.1590/S0034-71082000000200006

Quinlan, J.R., 1986. Induction of decision trees. Machine Learning, v. 1 (1), 81-106. https://doi.org/10.1007/BF00116251

Riahi, I., 2020. How hominin dispersals and megafaunal extinctions influenced the birth of agriculture. Journal of Economic Behavior & Organization, v. 175, 227-250. https://doi.org/10.1016/j.jebo.2020.03.025

Singh, G.; Reynolds, C.; Byrne, M.; Rosman, B.A., 2020. A remote sensing method to monitor water, aquatic vegetation, and invasive water hyacinth at national extents. Remote Sensing, v. 12 (24), 4021. https://doi.org/10.3390/rs12244021

Song, W.; Song, W.; Gu, H.; Li, F., 2020. Progress in the remote sensing monitoring of the ecological environment in mining areas. International Journal of Environmental Research and Public Health, v. 17 (6), 1846. https://doi.org/10.3390/ijerph17061846

Souza, J.F.S.; Neto, A.R.; Luque, S.P.; Gosset, M., 2024. Application of Sentinel-2 Level-2A images for monitoring water surface in reservoirs in the semiarid region of Pernambuco-Brazil. Revista Brasileira de Ciências Ambientais (RBCIAMB), v. 56 (1), e1927. https://doi.org/10.5327/Z2176-94781927

Thamaga, K.H.; Dube, T., 2018. Remote sensing of invasive water hyacinth (*Eichhornia crassipes*): a review on applications and challenges. Remote Sensing Applications: Society and Environment, v. 10, 36-46. https://doi.org/10.1016/j.rsase.2018.02.005

U.S. Geological Survey (USGS), 2022. Landsat missions. https://www.usgs.gov/landsat

Villamagna, A.M.; Murphy, B.R., 2010. Ecological and socioeconomic impacts of invasive water hyacinth (*Eichhornia crassipes*): a review. Freshwater Biology, v. 55, 282-298. https://doi.org/10.1111/j.1365-2427.2009.02294.x

Yang, C.; Clarke, K.; Shekhar, S.; Tao, C.V., 2020. Big spatiotemporal data analytics: a research and innovation frontier. International Journal of Geographical Information Science, v. 34 (6), 1075-1088. https://doi.org/10.1080/13658816.2019.1698743