

Phenological and Species-Level Classification of Aquatic Invasive Plants Using UAV Multispectral Imagery and Machine Learning

Daniel H. C. Salim¹, Caio C. S. Mello², Gabriel Pereira³, Raian V. Maretto⁴, Frederico Santos Machado⁵, Camila C. Amorim⁶

¹ Dept. of Sanitary and Environmental Engineering, Universidade Federal de Minas Gerais, Brazil – danielhcsalim@ufmg.br

² Dept. of Sanitary and Environmental Engineering, Universidade Federal de Minas Gerais, Brazil – caiocsm@ufmg.br

³ Department of Geosciences, Universidade Federal de São João del-Rei, Brazil - pereira@ufs.edu.br

⁴ Faculty of Geo-Information Science and Earth Observation, University of Twente, the Netherlands - r.v.maretto@utwente.nl

⁵ Petrobras Research Center, Petróleo Brasileiro S.A. (Petrobras), Brazil - fredericomachado@petrobras.com.br

⁶ Dept. of Sanitary and Environmental Engineering, Universidade Federal de Minas Gerais, Brazil - camila@desa.ufmg.br

Keywords: Remote sensing, macrophytes, *Eichhornia crassipes*, invasive species, reservoir

Abstract:

Monitoring aquatic invasive plant species (AIPs) and their phenological stages remains a challenge in complex freshwater environments. This study evaluates the potential of UAV-based multispectral imagery and machine learning for classifying six vegetation classes in a tropical urban reservoir composed of three phenological stages of *Eichhornia crassipes* and *Brachiaria subquadriflora*, *Pistia stratiotes*, and *Typha domingensis* species distribution. UAV flights were conducted on three dates using the MicaSense RedEdge-Dual sensor. A two-step principal component analysis (PCA) was used to select spectral bands and derive Normalized Difference and Ratio indices, aiming to reduce redundancy and assess their usefulness in classification. Three classifiers—Linear Discriminant Analysis (LDA), Random Forest (RF), and Support Vector Machine with RBF kernel (SVM-RBF)—were tested using 5-fold cross-validation. RF and SVM-RBF achieved the highest accuracies, ranging from 0.71 to 0.84, while LDA presented the lowest accuracy, between 0.63 and 0.82. Including spectral indices yielded only marginal improvements and did not consistently enhance classification performance, particularly when using more robust algorithms like RF and SVM-RBF, indicating that the ten original spectral bands are adequate to capture the key spectral distinctions in most cases. Classification performance was more consistent for *Brachiaria subquadriflora* and *Pistia stratiotes*, while considerable confusion was observed between *Typha domingensis* and the phenological stages of *Eichhornia crassipes*, likely due to spectral similarity. Overall, model selection had a higher influence on performance than feature augmentation. Future studies should explore spatial-textural features and sensor fusion to improve the generalization of AIP monitoring systems.

1. Introduction

The rapid proliferation of aquatic invasive plants (IAP), such as *Eichhornia crassipes* (water hyacinth), has significantly transformed the ecological balance of eutrophic reservoirs, resulting in biodiversity loss, deterioration of water quality, reduction of water surface area, and disruption of multiple water uses (Hestir *et al.*, 2008; Wu and Ding, 2020). These impacts are especially critical in tropical and subtropical urban reservoirs, where high nutrient loads and low water circulation rates create favorable conditions for uncontrolled macrophyte growth. Monitoring the spatial distribution, species composition, and phenological stages of these plants is crucial for understanding ecosystem dynamics, assessing pollution levels, and guiding control strategies and ecological restoration efforts (Wu and Ding, 2020).

Remote sensing has become an essential tool in the detection and monitoring of IAPs. Satellite data, such as from the Sentinel or Landsat programs, provide historical and regional-scale analysis, but their spatial resolution often limits their effectiveness in heterogeneous environments or in detecting early-stage infestations. UAVs equipped with multispectral sensors have emerged as a powerful complementary platform, offering sensor customization, sub-meter spatial resolution, and on-demand flight flexibility. These characteristics enable the precise mapping of aquatic vegetation, even in narrow or fragmented water bodies, and make UAVs effective for distinguishing species and phenological stages (Pádua *et al.*, 2022; Mouta *et al.*, 2023).

Despite these technological advances, challenges remain in accurately discriminating IAP species and their phenological stages, mainly due to their high spectral similarity, environmental variability, and the structural complexity of vegetation. Previous studies have shown that integrating spectral indices, such as NDVI or band ratios, alongside in-band reflectance data can sometimes improve classification performance, although results vary depending on sensor type, season and classification algorithm (Ade *et al.*, 2022; Mu *et al.*, 2023).

Machine learning techniques have increasingly been used to address these challenges by modeling complex, non-linear relationships in multispectral data. Among the commonly adopted classifiers, Linear Discriminant Analysis (LDA), Random Forest (RF), and Support Vector Machine with a radial basis function kernel (SVM-RBF) are promising techniques for vegetation mapping tasks. While LDA offers simplicity and interpretability, RF and SVM-RBF typically achieve higher accuracy in high-dimensional datasets due to their ability to capture intricate patterns. Evaluating these classifiers in the context of multitemporal UAV datasets and exploring their derived spectral indices can provide important insights into the robustness and generalizability of IAP classification models.

This study aims to improve the classification of IPA species and their phenological stages using high-resolution multispectral imagery acquired by UAVs. We evaluate and compare the performance of three machine learning classifiers, LDA, RF and SVM-RBF, in distinguishing six vegetation classes, including three growth stages of *Eichhornia crassipes*. To enhance model inputs and reduce feature redundancy, we implement a two-step

principal component analysis (PCA) for selecting the most informative spectral bands and derived indices (Normalized Difference and Ratio). The study also investigates whether the addition of these spectral indices improves classification performance across multiple UAV acquisition flights, and assesses the potential and limitations of each classifier in handling both species-level and intra-species phenological variation.

2. Methodology

To support this objective, three UAV campaigns were conducted at the Ibirité Reservoir (Minas Gerais, Brazil) on August 23, September 17, and October 2, 2024, using the MicaSense RedEdge Dual-P sensor, which captures data in ten multispectral bands at Blue-444, Blue-475, Green-531, Green-560, Red-650, Red-668, Red Edge-705, Red Edge-717, Red Edge-740, and Near Infrared-842 nm. These spectral bands are renamed as band_1 to band_10, respectively.

The three UAV campaigns were strategically timed to capture phenological variation in aquatic plant development, representing late winter (August), early spring (September), and mid-spring (October) conditions. According to Ghoussein *et al.* (2019), this period corresponds to the peak seasonal expansion of floating macrophytes, such as *Eichhornia crassipes*, when the species exhibit strong phenological expression and differences in spectral behaviour become more apparent. The dry-season stability and high vegetation cover observed during these months provide optimal conditions for class discrimination in multispectral imagery.

All flights were calibrated with a reflectance panel and were synchronized with Sentinel-2 overpasses, enabling future synergy between aerial and satellite data. The images were processed using Agisoft Metashape to generate high-resolution orthomosaics with a spatial resolution of 6 cm or less. Figure 1 illustrates an example of the orthomosaics. From these data, spectral reflectance data were systematically extracted from georeferenced polygons representing different vegetation classes.

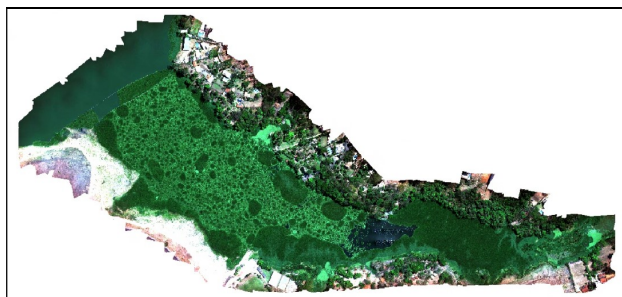


Figure 1. UAV orthomosaic of the Ibirité Reservoir (October 2, 2024).

For each UAV acquisition date, vegetation samples were extracted using labelled shapefiles containing georeferenced polygons that delineated homogeneous patches of known species and phenological stages. These shapefiles were used to spatially mask the multispectral UAV orthomosaics, allowing the extraction of pixel-level reflectance data exclusively within each polygon. Each polygon contributed individual pixel spectra to the dataset, rather than averaged values, resulting in a per-pixel classification approach.

To maintain class balance and avoid model bias, a maximum of 3,000 valid samples per vegetation class were randomly selected from each date. The final dataset comprised six classes: three phenological stages of *Eichhornia crassipes*, young floating, adult floating, and adult rooted (fixed in margins), along with *Brachiaria subquadrifida*, *Typha domingensis*, and *Pistia stratiotes*. Each sampled pixel was characterized by a ten-dimensional spectral vector corresponding to the UAV sensor's bands and a categorical class label.

Individual datasets were compiled for each UAV flight date and subsequently merged into a unified multitemporal database to support comparative and temporal classification analyses. For model evaluation, the full dataset was randomly partitioned into five stratified folds, ensuring a consistent 80% training and 20% validation split per fold across all classes.

The labelled training data were used to generate spectral signatures for each vegetation class across the three acquisition dates, highlighting both interclass separability and phenological variation over time. These spectral profiles, derived from pixel-level reflectance values, are presented in Figure 2. To provide spatial context for the labelled areas used in sample extraction, Figure 3 displays four zoomed-in orthomosaics with overlaid polygons, each representing a different invasive species or phenological stage in the reservoir. Complementing this, Figure 4 includes field photographs documenting the surveyed aquatic invasive plants during the UAV campaigns, offering a visual reference for the vegetation classes analysed in this study.

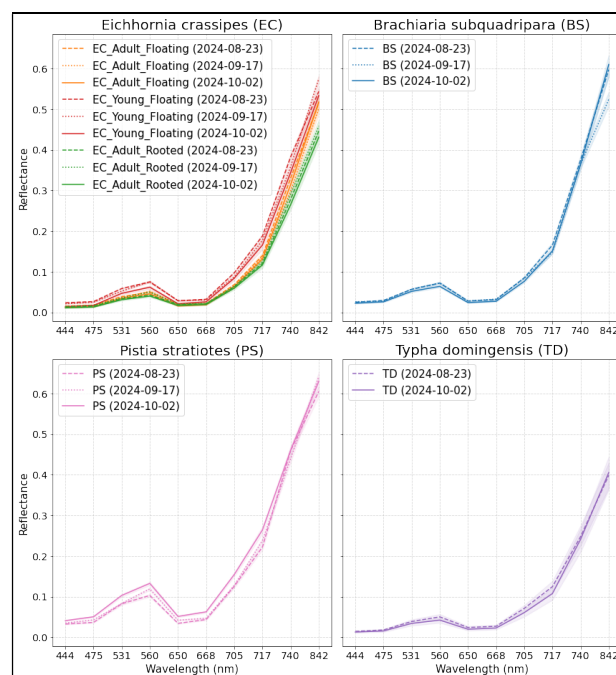


Figure 2. Spectral signatures of AIP classes derived from labels collected on each day of the UAV campaigns. *Typha domingensis* was not present in the mapped area on September 17, 2024.

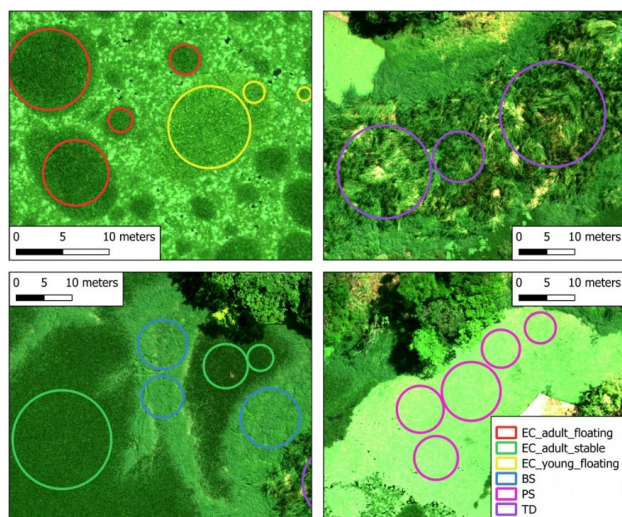


Figure 3. Labeled training polygons for aquatic vegetation classes used in the classification, delineating homogeneous patches of six vegetation classes.

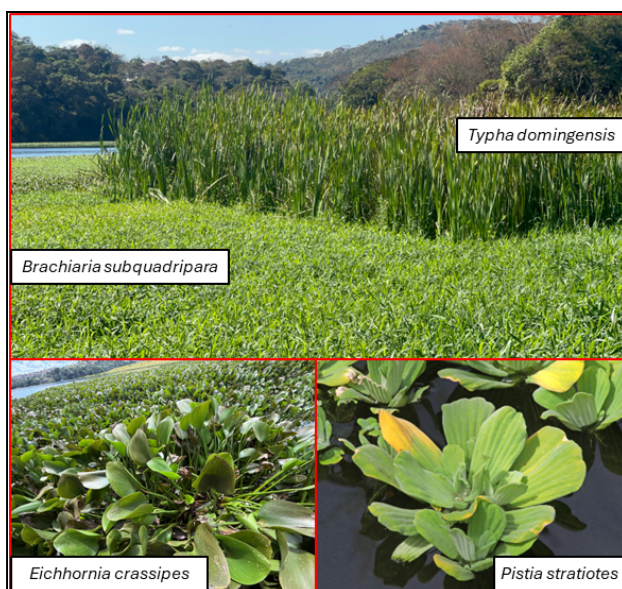


Figure 4. Aquatic invasive plant species observed in the study area.

To enhance the relevance of input variables, reduce multicollinearity, and improve the spectral separability of aquatic vegetation classes, a two-stage feature selection strategy was implemented based on unsupervised principal component analysis (PCA). This procedure was applied independently to each acquisition date, as well as to the combined multitemporal dataset. In the first stage, PCA was conducted on the full set of ten spectral bands acquired by the UAV sensor. Bands with the highest absolute loadings on the leading principal components were prioritized for selection, as they contributed most significantly to the variance structure of the data. To ensure the spectral diversity of the selected bands, pairwise cosine similarity was calculated between loading vectors, and bands with high spectral redundancy (cosine similarity > 0.90) were excluded.

Subsequently, all pairwise combinations of the selected, spectrally distinct bands were used to compute two types of

vegetation indices—Normalized Difference (ND) and Ratio—which are widely used to capture relative contrast and proportional relationships in vegetation reflectance. A second, independent PCA was then performed exclusively on the full set of derived indices. Importantly, this approach to index selection was entirely unsupervised and did not incorporate class labels. The same criteria were applied to retain only the most informative and non-redundant indices, yielding a refined set of variables for classification. This two-step unsupervised process resulted in a compact feature set composed of selected spectral bands and vegetation indices optimized for spectral variance while avoiding collinearity.

To evaluate the impact of feature selection on classification accuracy, two modelling strategies were employed. The first configuration (dataset B) used the full set of ten original spectral bands without any selection. The second configuration (dataset B+I) incorporated the reduced set of spectral bands and indices obtained through the unsupervised PCA-based procedure. All features were standardized using z-score normalization before classification. Three machine learning algorithms, LDA, RF and SVM-RBF were applied to both configurations (Thamaga and Dube, 2019; Bayable *et al.*, 2023).

This experimental design enabled a direct evaluation of how the PCA-based, unsupervised feature selection pipeline influenced the performance of each classifier. By comparing models trained on the original spectral bands (B) with those using the optimized combination of selected bands and vegetation indices (B+I), we systematically assessed the added value of unsupervised index selection for species-level classification of aquatic invasive plants in this study.

Mean performance metrics for the five stratified folds, including overall accuracy, precision, recall, and F1-score, were computed for each classifier and modelling strategy. Confusion matrices were also analysed to evaluate classification consistency across species and phenological stages. This comparative framework enabled a direct examination of how the inclusion of selected spectral indices affects model performance relative to using the original spectral bands alone.

3. Results:

3.1 PCA-based unsupervised feature selection

To reduce feature dimensionality and enhance class separability, a two-level PCA was applied: first to the original spectral bands and then to the derived spectral indices (Normalized Difference and Ratio). The first PCA, conducted on the ten-band reflectance data, identified bands with the highest variance contribution and lowest angular redundancy. As illustrated in Figure 5, which shows the PCA biplot for the combined dataset (ALL), selected bands such as band_3, band_5, band_8, and band_10 were the most informative and spectrally distinct, forming the basis for the generation of pairwise spectral indices.

The second PCA, applied to ND and Ratio indices derived from these bands, further reduced dimensionality by identifying a subset of indices with high explanatory power and minimal collinearity. The biplot shown in Figure 6 presents the results of this analysis for the ALL dataset, highlighting indices such as ND_B5_B3 – Normalized Green Red Difference Index (Tucker, 1979), ND_B5_B8, ND_B10_B3 – Green Normalized Difference Vegetation Index (Gitelson *et al.*, 1996), and R_B5_B8 – Ratio Vegetation Index (Birth and McVey, 1968), as the most relevant for class separation. These indices were

particularly effective at differentiating among phenological stages of *Eichhornia crassipes* and distinguishing other species like *Typha domingensis* and *Brachiaria subquadriflora*.

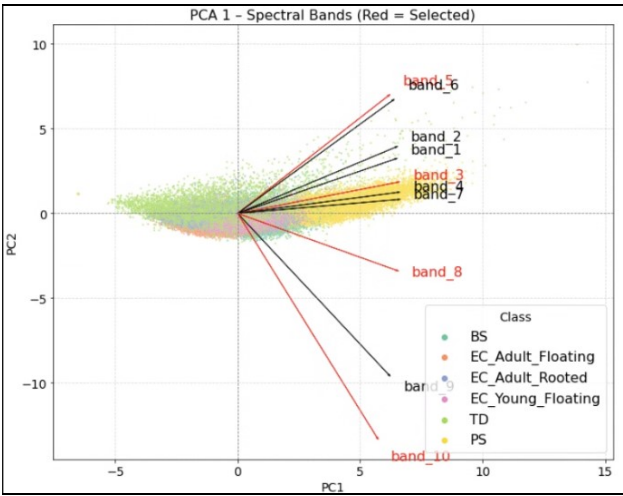


Figure 5. PCA biplot of the original spectral bands (band_1 to band_10) for the combined (ALL) dataset. Vectors represent the contribution of each band to the first two principal components, supporting the selection of bands 3, 5, 8 and 10 for spectral index generation.

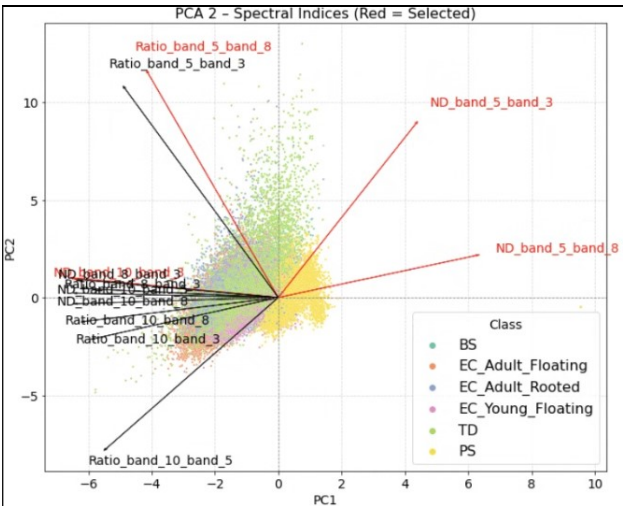


Figure 6. PCA biplot of the derived spectral indices based on selected bands from the first PCA. Highlighted indices (e.g., ND_B5_B3, ND_B5_B8, ND_B10_B3, and R_B5_B8) showed the highest variance contribution and were most effective for class discrimination.

Table 1 summarizes the spectral indices selected through PCA for each acquisition date and for the combined (ALL) multitemporal dataset. While the specific indices varied across dates—likely due to phenological changes and environmental conditions—certain spectral bands were consistently prominent. band_5 was the most frequently selected variable in this procedure, followed by band_10, band_3, and band_8. Additionally, band_6 and band_7 also appeared in multiple indices, reinforcing their importance for capturing subtle spectral variations relevant to vegetation class discrimination. The frequent recurrence of these bands underscores their robustness and relevance in multitemporal classification workflows.

Date	PCA Selected Indices
ALL	R B5 B8, ND B5 B3, ND B5 B8, ND B10 B3
23/08/2024	R B5 B7, R B10 B7, ND B10 B5, ND B8 B7
17/09/2024	ND B6 B3, R B6 B8, R B10 B3, R B10 B6
02/10/2024	R B5 B8, R B10 B5, ND B10 B3

Table 1. PCA-selected spectral indices for each acquisition date and the combined (ALL) dataset.

3.2 Classifier performance assessment

The 5-fold cross-validation classification performance across different machine learning algorithms and feature configurations revealed consistent trends throughout the multitemporal datasets. Both SVM-RBF and RF classifiers consistently achieved the highest overall accuracy and macro-averaged precision, recall, and F1-score across all acquisition dates. In contrast, Linear Discriminant Analysis (LDA) systematically underperformed, particularly in the October 2, 2024 dataset, where its accuracy (0.630) and F1-score (0.660) were substantially lower than those of RF (accuracy: 0.718; F1-score: 0.735) and SVM (accuracy: 0.720; F1-score: 0.738). These results highlight the advantage of non-linear and ensemble-based methods in modeling the complex spectral responses and high inter-class variability associated with aquatic invasive plants. For comparison, previous studies using Sentinel-2 data and RF, SVM for IAPs classification achieved overall accuracies above 85% (Pádua *et al.*, 2022; Bayable *et al.*, 2023; Mqingwana *et al.*, 2024).

Although including derivative spectral indices (B+I) is generally expected to enhance classification performance, especially in *Eichhornia crassipes* mapping tasks (Thamaga and Dube, 2019), our results showed inconsistent improvements. In several cases, particularly for RF and SVM-RBF, models using only the original spectral bands (B) performed comparably to, or even slightly better than, using a complex dataset. This suggests that in high spatial and spectral resolution, the raw spectral information may already capture the key variability needed for class discrimination.

This trend suggests that these algorithms are inherently robust to input dimensionality and may already capture the key spectral distinctions using raw reflectance data. An exception was observed in the LDA classifier on the October 2 dataset, where the inclusion of indices improved classification accuracy from 0.630 to 0.685 and F1-score from 0.660 to 0.700, indicating that feature enhancement may be more beneficial for linear classifiers when class separability is limited.

As shown in Table 2, classification performance on the full multitemporal dataset (ALL) remained stable for RF and SVM, with overall accuracies ranging from 0.76 to 0.78 regardless of whether indices were included. LDA, however, continued to perform poorly under the multitemporal configuration, reaffirming its limitations in capturing spectral variability across acquisition dates. Although including derivative spectral indices led to minor performance gains in specific cases, most notably with LDA, their overall contribution was limited when compared to the effect of the classification algorithm itself. Comparative results based on the F1-score indicate that model choice had a far greater influence on classification success than feature augmentation alone.

Date	Model	Features	Accuracy	Precision	Recall	F1-Score
23/08/2024	LDA	B+I	0.764	0.764	0.764	0.764
23/08/2024	LDA	B	0.766	0.761	0.762	0.762
23/08/2024	RF	B+I	0.804	0.803	0.804	0.803
23/08/2024	RF	B	0.807	0.807	0.806	0.806
23/08/2024	SVM (RBF)	B+I	0.821	0.820	0.821	0.820
23/08/2024	SVM (RBF)	B	0.822	0.823	0.821	0.821
17/09/2024	LDA	B+I	0.811	0.808	0.811	0.808
17/09/2024	LDA	B	0.818	0.818	0.817	0.817
17/09/2024	RF	B+I	0.826	0.824	0.826	0.824
17/09/2024	RF	B	0.827	0.825	0.825	0.825
17/09/2024	SVM (RBF)	B+I	0.840	0.839	0.840	0.839
17/09/2024	SVM (RBF)	B	0.839	0.837	0.838	0.838
02/10/2024	LDA	B+I	0.685	0.700	0.685	0.700
02/10/2024	LDA	B	0.630	0.729	0.660	0.660
02/10/2024	RF	B+I	0.714	0.731	0.714	0.731
02/10/2024	RF	B	0.718	0.808	0.735	0.735
02/10/2024	SVM (RBF)	B+I	0.719	0.737	0.719	0.737
02/10/2024	SVM (RBF)	B	0.720	0.810	0.738	0.738
ALL	LDA	B+I	0.701	0.766	0.701	0.701
ALL	LDA	B	0.714	0.736	0.708	0.708
ALL	RF	B+I	0.780	0.780	0.752	0.752
ALL	RF	B	0.763	0.782	0.759	0.759
ALL	SVM (RBF)	B+I	0.761	0.780	0.758	0.758
ALL	SVM (RBF)	B	0.762	0.780	0.757	0.757

Table 2. Performance Metrics of LDA, Random Forest, and SVM (RBF) Models Across Dates and Feature Sets

These findings emphasize that algorithm selection is more essential than feature engineering when classifying AIP species using UAV-based multispectral imagery. Ensemble methods like Random Forest and kernel-based approaches such as SVM (RBF) consistently delivered higher accuracy and F1-scores across dates and configurations, demonstrating strong performance and adaptability over time. This makes them highly suitable for operational monitoring of aquatic vegetation in dynamic environments.

Regarding the classification of the dataset ALL with B+I features, the confusion matrices for LDA (Figure 7), RF (Figure 8), and SVM-RBF (Figure 9) reveal important patterns about the strengths and limitations of each classifier in handling both inter-species and intra-species variation in AIP.

Species-level differentiation was generally strong across all models, particularly for *Brachiaria subquadriflora*, *Eichhornia crassipes*, and *Pistia stratiotes*. These three species displayed consistently high true positive rates with limited confusion among one another, suggesting that their spectral signature variations are sufficiently distinct to be reliably detected by all classifiers. In contrast, *Typha domingensis* proved more difficult to classify consistently. All three models showed confusion between *Typha domingensis* and other classes, particularly with *Eichhornia crassipes*, which likely reflects the higher spectral variability of *Typha*.

Typha is a tall and emergent macrophyte, with upright stems and denser vertical structure, and exhibits a complex interaction with light, including variable shadowing, reflectance anisotropy, and increased within-class heterogeneity (Al-Sodany *et al.*, 2022). Additionally, *Typha* tends to remain physiologically active over longer periods, introducing temporal variation in chlorophyll content, canopy moisture, and structural traits. These ecological features result in a broader spectral footprint across time and space, increasing the likelihood of spectral overlap with other species. Also, this variability reduced its separability under both linear and non-linear classification schemes.

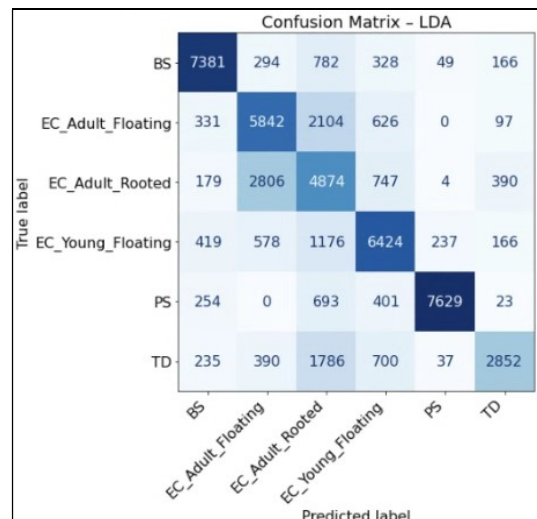


Figure 7. Confusion matrix of LDA classification using B+I features (Dataset: ALL).

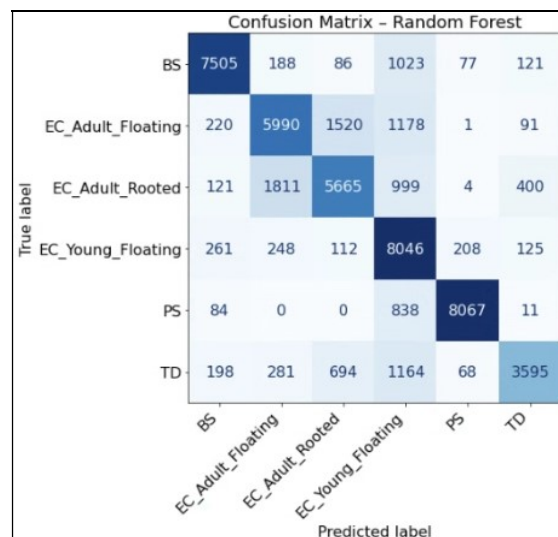


Figure 8. Confusion matrix of RF classification using B+I features (Dataset: ALL).

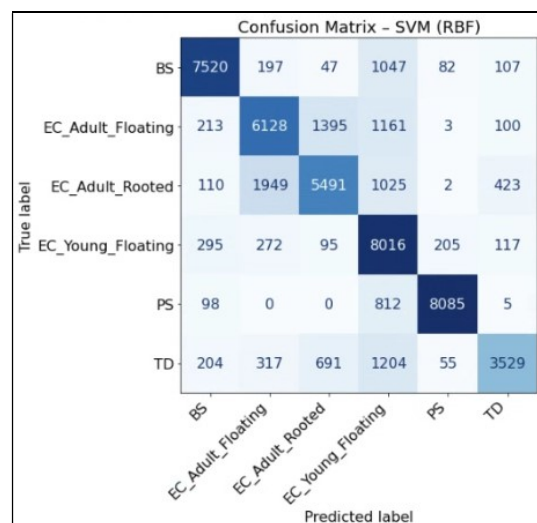


Figure 9. Confusion matrix of SVM-RBF classification using B+I features (Dataset: ALL).

The confusion observed between adult-floating and adult-rooted *Eichhornia crassipes* in the classification results is likely due to their similar reflectance profiles in the red-edge and near-infrared regions. While floating adults tend to maintain higher reflectance in these wavelengths due to healthier, more hydrated leaves, long-term adult-rooted plants at the margins often exhibit increased senescence, homogeneous and denser canopy structures—factors that reduce red-edge and NIR reflectance. Ghoussein et al. (2023) demonstrated that *Eichhornia crassipes* with signs of aging or mixed physiological states (e.g., yellowing or flowering individuals) presented lower reflectance in NIR bands compared to vigorous, green monospecific stands.

In contrast, young plants are more easily distinguishable from adult individuals due to their higher reflectance in the visible range. This spectral behavior is linked to their lower chlorophyll content, which results in reduced absorption and consequently greater reflectance in the visible wavelengths. Sims and Gamon (2002) found this pattern to be consistent across species and developmental stages, highlighting that young leaves, with underdeveloped pigment structures, exhibit distinct optical properties compared to mature leaves. As a result, phenological differences between young and adult plants generate a more pronounced spectral separation than the structural differences between adult floating and rooted forms, whose divergence is primarily constrained to the NIR and red-edge domains.

The feature importance analysis from the Random Forest classifier trained on the full multitemporal dataset (ALL, B+I) revealed a clear dominance of raw spectral bands over derivative indices (Table 3). The top-ranked features included band_4 (Green, 560 nm), band_3 (Green, 531 nm), and band_7 (Red Edge, 705 nm), followed by band_1 (Blue, 444 nm) and band_10 (NIR, 842 nm). These wavelengths are strongly associated with pigment content, internal canopy structure, and water status—key traits for distinguishing aquatic invasive species under varying phenological conditions.

This ranking aligns with Hestir et al. (2008), who emphasized the importance of red-edge and NIR reflectance in capturing the structural complexity and physiological variation within *Eichhornia crassipes* mats, particularly where flowering, senescence, and active growth coexist. Notably, ND band_10 band_3, the GNDVI, was among the highest-ranked indices, reflecting its sensitivity to chlorophyll concentration and biomass. Mucheye et al. (2022) reported strong correlations between GNDVI and chlorophyll content in *E. crassipes*, reinforcing its use for classification of phenological stages.

Band	Importance Score
band_4 (Green-560)	0.1077
band_3 (Green-531)	0.1040
band_7 (Red Edge 705)	0.0947
band_1 (Blue-444)	0.0929
band_10 (Near Infrared-842)	0.0855
band_8 (Red Edge-717)	0.0852
band_2 (Blue-475)	0.0796
ND_band_10_band_3	0.0700
band_9 (Red Edge-740)	0.0670
ND_band_5_band_3	0.0511
band_6 (Red-668)	0.0507
ND_band_5_band_8	0.0405
Ratio_band_5_band_8	0.0384
band_5 (Red-650)	0.0349

Table 3. Top features ranked by RF Importance for the ALL Dataset using B+I features.

Our study builds upon recent advances demonstrating the effectiveness of RF and SVM in remote sensing-based mapping of *Eichhornia crassipes*. Bayable et al. (2023), working with Sentinel-2 and Landsat-8 imagery in Lake Tana (Ethiopia), reported overall accuracies exceeding 0.9 using RF and SVM for distinguishing *E. crassipes* from surrounding land cover types. Similarly, Pádua et al. (2022) applied these algorithms to UAV multispectral imagery over the Mondego River basin (Portugal) in a binary classification of *E. crassipes* versus open water, achieving accuracies of 0.94 with RF and 0.87 with SVM. While these results confirm the effectiveness of these algorithms, they were obtained under relatively simplified classification schemes, with greater spectral separation between classes and limited intra-class variability. In contrast, our study applied RF and SVM-RBF to a more complex and ecologically realistic scenario involving six vegetation classes within a heterogeneous urban reservoir. Despite the higher spectral similarity and structural variability among classes, our models achieved overall accuracies ranging from 0.71 to 0.84. These findings reaffirm the robustness of RF and SVM in aquatic vegetation mapping, particularly when applied to fine-resolution UAV data, and underscore their suitability for operational monitoring in more challenging classification contexts.

3.3 LDA Projection and the effect of spectral indices

Regarding the LDA classification of the October 2 dataset, where the inclusion of indices improved classification accuracy and F-1 score, Figure 10 illustrates the corresponding LDA projections in two scenarios: using only the original spectral bands (top panel) and using both bands and derivative indices (bottom panel). Although the overall distinction between the two scenarios is not remarkable, the most notable improvement is observed in the separation between *Eichhornia crassipes* adult rooted (green points) and adult floating (yellow points), which show reduced overlap and greater distance between their centroids in the augmented feature space (Bands + Indices). This suggests that the inclusion of spectral indices contributed to enhancing the discriminability between these two phenological forms.

Furthermore, the *Eichhornia crassipes* young floating (red points) and *Typha domingensis* (purple points) classes became more adjacent in the LDA projection with indices, indicating a potential trade-off where the additional features did not consistently improve class separability across all groups. A further observation is the increased isolation of the *Brachiaria subquadriflora* class (blue points) when indices were included, as its centroid moved farther away from the other groups, suggesting improved distinction from spectrally similar vegetation.

These subtle shifts in the LDA space demonstrate that the inclusion of spectral indices can enhance linear discriminant performance for certain class pairs, particularly when original bands alone do not fully capture the spectral variability associated with structural and physiological plant traits. Despite its simplicity and interpretability, LDA's limited separation reinforces the need for non-linear models like RF and SVM-RBF to effectively capture complex spectral patterns inherent to IAPs (Pádua et al., 2022; Bayable et al., 2023).

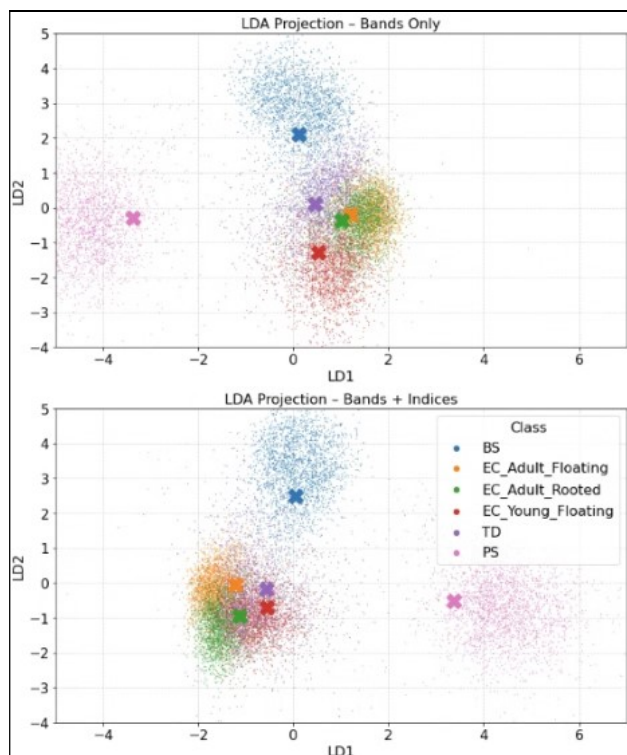


Figure 10. LDA projection of the October 2 dataset using B (Top) and B+I (Bottom) for IAP classification.

4. Conclusions

This study advanced the classification of aquatic invasive plant species and their phenological stages using high-resolution UAV multispectral imagery combined with machine learning algorithms. A two-level PCA-based dimensionality reduction was applied to identify the most informative spectral bands and construct meaningful Normalized Difference and Ratio indices. Although the selected indices visually separated the vegetation classes in PCA space, suggesting potential for improved class separability, they did not consistently enhance the performance of classification models compared to using the full set of ten original spectral bands. This indicates that raw reflectance data already captures the key spectral variation necessary for accurate classification in most cases.

RF and SVM-RBF achieved higher performance metrics than LDA across all acquisition dates and feature configurations. These results highlight the advantages of ensemble and kernel-based methods in modelling the non-linear spectral patterns associated with aquatic vegetation. Among the species analysed, *Brachiaria subquadripara* and *Pistia stratiotes* exhibited the highest spectral distinction from *Eichhornia crassipes*, enabling more reliable classification. In contrast, *Typha domingensis* showed significant spectral overlap with all phenological stages of *Eichhornia*, leading to frequent misclassifications. These difficulties are likely related to *Typha*'s taller emergent structure and high intra-class variability.

Future studies should focus on integrating object-based image analysis (GEOBIA) to incorporate spatial, morphological, and textural information, which can enhance the ability to distinguish spectrally similar classes. The adoption of deep learning methods, particularly Convolutional Neural Networks (CNNs), offers promising potential for learning complex spectral-spatial features directly from UAV imagery.

Additionally, the fusion of UAV and satellite datasets, such as Sentinel-2, can increase the temporal and spatial scalability of classification systems, making them more suitable for continuous monitoring. Spectro-temporal modelling and the inclusion of structural metrics are also recommended to improve robustness over time and better capture ecological dynamics.

Acknowledgement

The authors acknowledge Petrobras/ANP for supporting the AQUASMART Research and Development Project, under which this study was conducted. We also thank the technical staff and colleagues involved in data collection and processing, as well as the Universidade Federal de Minas Gerais (UFMG) for equipment and logistical support. Additional support was provided by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

References

- Ade, C. *et al.* (2022) 'Genus-Level Mapping of Invasive Floating Aquatic Vegetation Using Sentinel-2 Satellite Remote Sensing', *Remote Sensing*, 14(13), pp. 1–20. Available at: <https://doi.org/10.3390/rs14133013>.
- Al-Sodany, Y.M. *et al.* (2022) 'Regression models to estimate accumulation capability of six metals by two macrophytes, typha domingensis and typha elephantina, grown in an arid climate in the mountainous region of taif, Saudi Arabia', *Sustainability (Switzerland)*, 14(1), pp. 1–22. Available at: <https://doi.org/10.3390/su14010001>.
- Bayable, G. *et al.* (2023) 'Detection of Water Hyacinth (*Eichhornia crassipes*) in Lake Tana, Ethiopia, Using Machine Learning Algorithms', *Water (Switzerland)*, 15(5). Available at: <https://doi.org/10.3390/w15050880>.
- Ghoussein, Y. *et al.* (2019) 'Multitemporal remote sensing based on an FVC reference period using sentinel-2 for monitoring *Eichhornia crassipes* on a Mediterranean river', *Remote Sensing*, 11(16). Available at: <https://doi.org/10.3390/rs11161856>.
- Hestir, E.L. *et al.* (2008) 'Identification of invasive vegetation using hyperspectral remote sensing in the California Delta ecosystem', *Remote Sensing of Environment*, 112(11), pp. 4034–4047. Available at: <https://doi.org/10.1016/j.rse.2008.01.022>.
- Mouta, N. *et al.* (2023) 'Sentinel-2 Time Series and Classifier Fusion to Map an Aquatic Invasive Plant Species along a River—The Case of Water-Hyacinth', *Remote Sensing*, 15(13). Available at: <https://doi.org/10.3390/rs15133248>.
- Mqingwana, P. *et al.* (2024) 'Monitoring and assessing the effectiveness of the biological control implemented to address the invasion of water hyacinth (*Eichhornia crassipes*) in Hartbeespoort Dam, South Africa', *Remote Sensing Applications: Society and Environment*, 36(July), p. 101295. Available at: <https://doi.org/10.1016/j.rsase.2024.101295>.
- Mu, S. *et al.* (2023) 'Identification for the species of aquatic higher plants in the Taihu Lake basin based on hyperspectral remote sensing', *Environmental Monitoring and Assessment*,

195(8). Available at: <https://doi.org/10.1007/s10661-023-11523-z>.

Pádua, L. *et al.* (2022) ‘Water Hyacinth (*Eichhornia crassipes*) Detection Using Coarse and High Resolution Multispectral Data’, *Drones*, 6(2), pp. 1–14. Available at: <https://doi.org/10.3390/drones6020047>.

Thamaga, K.H. and Dube, T. (2019) ‘Understanding seasonal dynamics of invasive water hyacinth (*Eichhornia crassipes*) in the Greater Letaba river system using Sentinel-2 satellite data’, *GIScience and Remote Sensing*, 56(8), pp. 1355–1377. Available at: <https://doi.org/10.1080/15481603.2019.1646988>.

Wu, H. and Ding, J. (2020) ‘Abiotic and Biotic Determinants of Plant Diversity in Aquatic Communities Invaded by Water Hyacinth [*Eichhornia crassipes* (Mart.) Solms]’, *Frontiers in Plant Science*, 11(August), pp. 1–11. Available at: <https://doi.org/10.3389/fpls.2020.01306>.