Detection and Mapping of Invasive Alien Plant Water Hyacinth using Satellite Imagery and Machine Learning (Case Study: Rawa Pening Lake, Indonesia)

A S Muammal[1], W Marsisno[1]

[1] Department of Statistical Computing, Politeknik Statistika STIS, Jakarta, Indonesia

*Corresponding author's email: waris@stis.ac.id

Abstract. Rawa Pening Lake, one of the 15 national priority lakes in Indonesia, faces a significant threat from invasive water hyacinth (Eichhornia crassipes). This plant once covered up to 70% of the lake's surface and continued to cause ecological and socio-economic impacts as of 2024, necessitating periodic monitoring to prevent future blooms. This study aimed to identify the optimal features to characterize water hyacinth, determine the most effective classification model, and map the plant's distribution. Adopting the CRISP-DM framework, the study utilized Sentinel-1 (radar) and Sentinel-2 (optical) satellite imagery with multispectral band features, radar bands, and composite indexes. Feature selection was performed using Jenks Natural Breaks, and classification modeling was conducted using Random Forest and Convolutional Neural Network (CNN). The results demonstrated that the CNN achieved higher accuracy in distinguishing among land cover classes. The final mapping identified water hyacinth covering 34,775 pixels, 32,627 pixels, and 34,175 pixels in June, July, and August, respectively. This approach offers a reliable method for periodic monitoring of water hyacinths in Rawa Pening Lake.

Keyword: deep learning, machine learning, remote sensing, water hyacinth.

1. Introduction

Water Hyacinth *(Eichhornia crassipes)* is an invasive aquatic plant considered one of the most aggressive and dominant invasive species in the world [1]. It is a highly productive photosynthetic plant capable of doubling its biomass within 6 to 14 days under suitable environmental conditions [2].Invasions of free-floating aquatic plants such as water hyacinths are considered a serious threat to biodiversity, deteriorating water quality, economic development, and human well-being. Once established, an invasion of aquatic plants can be difficult to control. Therefore, there is a need for regular monitoring of species that are included in invasive plants in a targeted and effective manner [3]. In response to the threat of invasive alien species, the United Nations (UN) has set Target 15.8 in the Sustainable Development Goals (SDGs), which focuses on efforts to prevent, mitigate, and control such species in terrestrial and aquatic ecosystems [4]. This global commitment is realized by Indonesia through the issuance of specific regulations, one of which is the Regulation of the Minister of Environment and Forestry (Permen LHK) Number P.94/MENLHK/SETJEN/KUM.1/12/2016

concerning Invasive Types. In the regulation, invasive species are defined as species that colonize habitats massively and can cause ecological, economic, and social losses [5].

Water hyacinths have spread to different parts of the world as seen in figure 9.

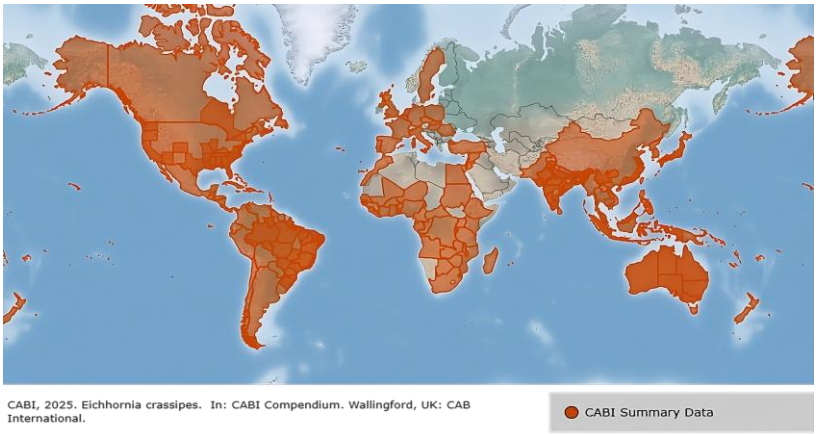


Figure 9. Worldwide distribution of water hyacinth [6].

Some parts of the world have experienced problems due to water hyacinth plants, the impact of which is not only detrimental to economic sectors, but also threatens the sustainability of ecosystems. In some African countries, such as Ethiopia and Nigeria, the spread of water hyacinths has caused significant disruption to socio-economic activities, including boat navigation, recreational access to waters, as well as fishing activities, and tourism [7].

The impact of water hyacinth is also felt in Indonesia, especially in aquatic ecosystems in the form of lakes, reservoirs, and rivers. Rawa Pening Lake, located in Semarang Regency, Central Java, is one of 15 national priority lakes that face major challenges related to water quality management and its ecosystems. Based on Presidential Regulation Number 60 of 2021, some of the problems faced in Rawa Pening are eutrophication and the dominance of aquatic weeds such as water hyacinths, which cover up to 70% of the lake area [8]. The presence of these water hyacinths has caused various problems, including lake siltation, disruption of water transportation, and degradation of scenic beauty, which reduces tourism aesthetics and inconveniences local anglers [9], [10]. This problem is exacerbated by Rawa Pening's water quality, which was classified as hypertrophic in 2022, with conditions ranging from meeting quality standards to being lightly polluted. [11].

Water hyacinths have excellent reproductive capabilities, especially in aquatic ecosystems that contain many nutrients such as nitrogen and phosphorus (eutrophic) [12]. This condition not only has an impact on the environment, but also on the economic and tourism sectors. According to [13], Rawa Pening is one of the leading artificial tourist destinations in the region, with the number of visits reaching 203,594 people in 2022 and 171,304 people in 2023, making it the second and third most visited artificial tourist attractions for two consecutive years. The high ecological pressure from the existence of water hyacinths can have implications for the decline in tourist attraction and the economy of the community around the lake.

Various efforts have been made to control the spread of water hyacinths in Indonesian waters, including in Rawa Pening Lake, by means of manual cleaning with the help of heavy equipment and human labor, the use of water hyacinths as organic fertilizers, and the use of biocontrol agents such as Grass Carp. Nonetheless, these measures are generally responsive and less effective in the long run. One of the preventive measures against the invasion of water hyacinths as an invasive plant is to conduct

periodic monitoring of aquatic ecosystems. However, monitoring of these invasive aquatic plants is still dominated by a field-based approach, which generally requires a lot of time, cost, and effort[14].

The limitations of field survey methods, such as high costs, time, and energy requirements, as well as limited spatial coverage, are significant obstacles in monitoring water hyacinths that are able to breed and move quickly due to currents and winds. Therefore, the use of remote sensing technology is an effective alternative, due to its ability to provide periodic and cost-effective observational data. Successful detection using satellite imagery relies heavily on the identification of the most effective distinguishing features to separate water hyacinths from other surrounding objects such as water, soil, and other vegetation. In optical data such as Sentinel-2, each object has a distinctive spectral mark based on how they reflect and absorb energy, so the challenge is to identify the best spectral features to accurately distinguish water hyacinths from other vegetation. However, because optical data is constrained by cloud cover, Sentinel-1's Synthetic Aperture Radar (SAR) data offers a complementary solution capable of operating in all weather conditions. Detection with SAR data is based on the identification of backscatter features, where calm water surfaces will appear dark, while rough expanses of water hyacinths will appear brighter as they cause volumetric scattering [15].

The use of satellite imagery in detecting water hyacinth has been widely carried out by several researchers [16]. The majority of the study used Landsat-8 and Sentinel-2 satellite imagery, which are known to have superior capabilities in observing vegetation and waters. To complement advances in remote sensing technology, machine learning algorithms such as Random Forest (RF) and deep learning algorithms such as Convolutional Neural Networks (CNN) have been used in extracting information from satellite imagery to detect plants, especially water hyacinths, more accurately. Random Forest (RF) is one of the most widely used algorithms due to its ability to handle multispectral data and generate high accuracy.

In addition, the deep learning approach using Convolutional Neural Networks (CNN) has proven to be superior, as in [17] and [18] which shows a spatio-temporal classification accuracy of 90% in detecting water hyacinths with Sentinel-2 imagery. CNN consistently outperforms traditional methods such as RF and K-Means, providing more reliable automated detection results that can be repeated every few days.

A literature review reveals a research gap in Indonesia regarding the detection of water hyacinth distribution using a synergistic integration of Sentinel-1 SAR and Sentinel-2 optical data with machine learning algorithms. Previous studies have demonstrated success using single-sensor approaches; for instance, research utilizing deep learning on Sentinel-2 imagery has achieved high spatio-temporal classification accuracy of 90%. However, these optical-only methods are inherently vulnerable to data loss from persistent cloud cover, a significant limitation in tropical regions. This study addresses that gap.

The theoretical superiority of combining Sentinel-1 and Sentinel-2 data lies in their complementary strengths, which create a more robust and consistent dataset than either could provide alone. Sentinel-2 offers high-resolution (up to 10 meters) multispectral optical data with a short temporal resolution (5 days), which is ideal for detailed spectral analysis of vegetation. Its primary weakness, however, is an inability to penetrate clouds. This is directly compensated for by Sentinel-1's Synthetic Aperture Radar (SAR) sensor, which provides all-weather imaging capabilities, ensuring data availability even in cloudy conditions. By fusing Sentinel-1's structural and textural information with Sentinel-2's detailed spectral information, this approach allows for continuous and reliable monitoring, overcoming the limitations of previous optical-based studies.

Based on this background, this study aims to develop and evaluate a model for accurately detecting and mapping water hyacinth distribution through a synergistic integration of radar (Sentinel-1) and optical (Sentinel-2) imagery with machine and deep learning algorithms. The main contribution of this research is the development of a comprehensive monitoring methodology that addresses a primary

limitation of previous optical-only studies: data loss due to frequent cloud cover in tropical regions. By integrating all-weather Sentinel-1 SAR imagery, this approach creates a more reliable and consistent monitoring framework suitable for an Indonesian priority lake.

This is achieved through three key objectives: first, systematically identifying the most effective combination of spectral, index, and radar features to uniquely characterize the target object; second, conducting a comparative analysis between Random Forest and Convolutional Neural Network (CNN) models to determine the most accurate and stable algorithm for this classification task ; and finally, applying the superior model to produce precise water hyacinth distribution maps for June, July, and August, providing a vital tool for ecological management and early warning systems against potential blooms in Lake Rawa Pening. The overall research framework is illustrated in Figure 10

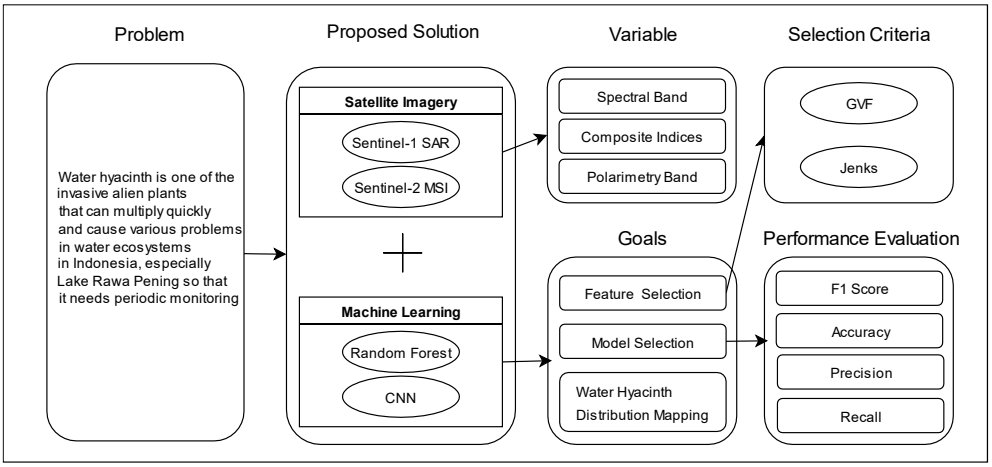


Figure 10. Proposed research framework.

2. Research Method

This study used the Cross-Industry Standard Process for Data Mining (CRISP-DM) approach, which provides a systematic and structured framework in the data analysis process. This approach was chosen because it has been proven to be able to improve work efficiency, process consistency, and support good documentation in data mining projects [19]. In addition, a study by [20] suggests that CRISP-DM is relevant for spatial modeling in land cover classification.

The stages of CRISP-DM implemented included: (1) Project Understanding, namely formulating the problem of the distribution of water hyacinths as an ecological and social threat in Rawa Pening Lake that requires periodic monitoring; (2) Data Understanding, which involved exploring the spectral and backscatter characteristics of Sentinel-1 SAR GRD and Sentinel-2 MSI Level-2A imagery obtained from the Google Earth Engine (GEE) platform. The primary acquisition time was focused on July 2024, with Sentinel-1 data from July 25 and Sentinel-2 from July 26, selected for their proximity to the aerial photography date (July 23, 2024) and a maximum cloud cover of 10%. For temporal analysis, supporting imagery from June and August 2024 was also used. This stage also included the creation of derivative variables such as NDVI, MNDWI, BSI, and FAI to understand the potential of distinguishing features between land cover classes; (3) Data Preparation, which encompasses all pre-processing activities before modeling. This stage began by establishing a Region of Interest (ROI) for the study area based on a shapefile from BBWS Pemali Juana, totaling 158,281 pixels at 10m resolution, which was used to clip all satellite imagery. The next step was data labeling, conducted in Google Earth Engine using point sampling. Guided by the protocol for land cover validation [21], which requires a minimum of 50 samples for studies with fewer than 12 classes, this research adopted a more robust sample size

of 200 samples for each of the six land cover classes: water hyacinth, floating net cages (KJA), water, other vegetation, bare soil, and built-up areas, resulting in 1200 total samples per image. A larger dataset was deliberately chosen to ensure sufficient data for training the deep learning model. CNNs, in particular, perform more effectively and generalize better with more extensive training examples, thereby increasing model stability and reducing the risk of overfitting. The collection of these samples was visually aided, with the reliability of the process varying by month. For the July imagery, labeling was conducted with high confidence using aerial photographs from BBWS Pemali Juana as the primary ground truth reference. For the June and August datasets, where drone imagery was not available, labeling was conducted through careful visual interpretation of Sentinel-2 true-color RGB composites. To account for potential object movement between the slightly different acquisition times of the Sentinel-1 and Sentinel-2 images, points were only placed where land cover was consistently identifiable across both the optical (RGB) and radar (backscatter) imagery.

Following data labeling, the preparation stage continued with feature identification using the Jenks Natural Breaks (JNB) method. This method was chosen for its strength as a data classification tool, ideal for an exploratory analysis aimed at understanding the inherent separability of land cover classes. The JNB algorithm optimally divides data into groups by minimizing the variance within each class while maximizing the variance between classes[22]. This approach helps visualize the natural data groupings to identify which features offer the clearest distinction between classes before modeling begins. The data were grouped into five levels (very low, low, medium, high, and very high), and the quality of this classification was evaluated using the Goodness of Variance Fit (GVF) value, where a value closer to 1 indicates a better distribution. Figure 11 is an example of how the JNB interval divides the distribution of data.

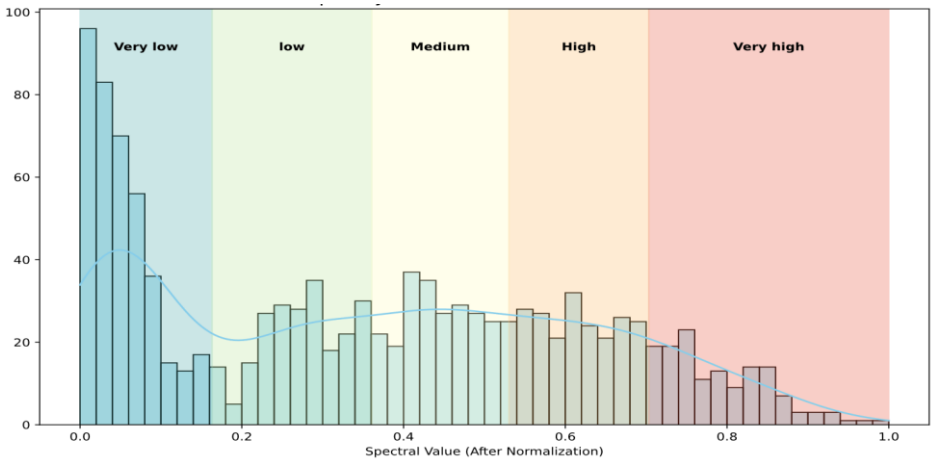


Figure 11. Example of JNB level division visualization.

The evaluation of the Jenks Natural Breaks (JNB) method was carried out with Goodness of Variance Fit to ensure the reliability of the class separation. The Goodness of Variance Fit value has a range between 0 to 1, where a value close to 1 indicates a better distribution outcome. After the Jenks Natural Breaks algorithm determined the optimal class boundaries, the next step was to convert this numerical classification into a visual heatmap representation. This process begins with the calculation of the median statistical value for each predictor variable (e.g., NDVI, NDWI, VV) on the set of pixels belonging to each land cover class. Separately, for each variable, the Jenks Natural Breaks algorithm was applied to the value distribution of all pixels on the sample dataset to determine five optimal quantitative levels, ranging from very low (Level 1) to very high (Level 5). The median grades of each class are then mapped into one of the five levels based on their range of values. The results of this mapping are then visualized as a heatmap, where rows represent variables, columns represent land cover

classes, and cell colors indicate their characteristic levels. Figure 12 shows the visualization process of the distribution of Jenks Natural Breaks levels into a heatmap.

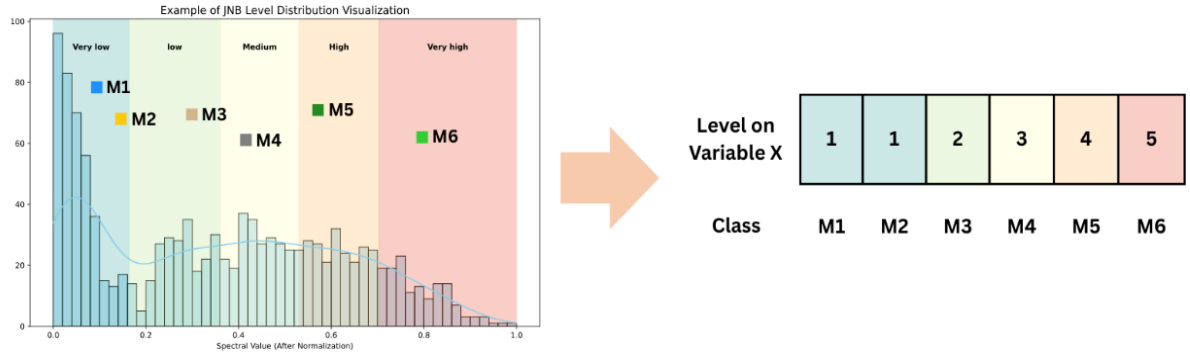


Figure 12. Example of the visualization process from Jenks Natural Breaks into a heatmap.

This heatmap analysis makes it possible to observe the differentiating power of each variable. Referring to the example of the visualization given, a variable can be considered a very strong distinguishing feature for the M6 class, since its median value is exclusively at Level 5 (Very High), so it is easily separated from all the other classes. The same variables also showed effectiveness in identifying classes M5 (Level 4) as well as M4 and M3 (Level 3, Level 2). In contrast, the variable proved ineffective in distinguishing between classes M1 and M2, as the median values of both classes fell into Level 1 (Low). This kind of interpretation suggests that the effectiveness of a variable is relative; it can be an excellent identifier for one class, but less useful for separating other classes that have similar statistical responses.; (4) Modeling, by applying Random Forest and one-dimensional Convolutional Neural Network (1D CNN) algorithms to a training dataset composed of 70% of the total samples, which was created using stratified sampling. In the deep learning approach, each pixel was treated as an input vector representing its spectral signature, and the 1D CNN architecture consisted of convolution, reshaping, fully connected, normalization, and activation layers as shown in Figure 13.

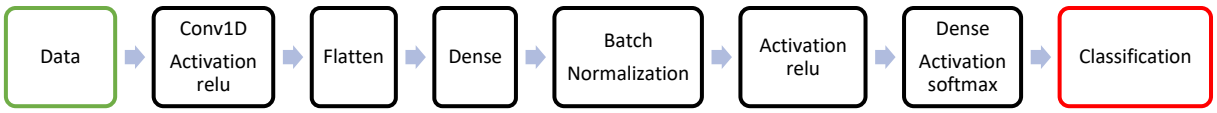


Figure 13. CNN architecture.

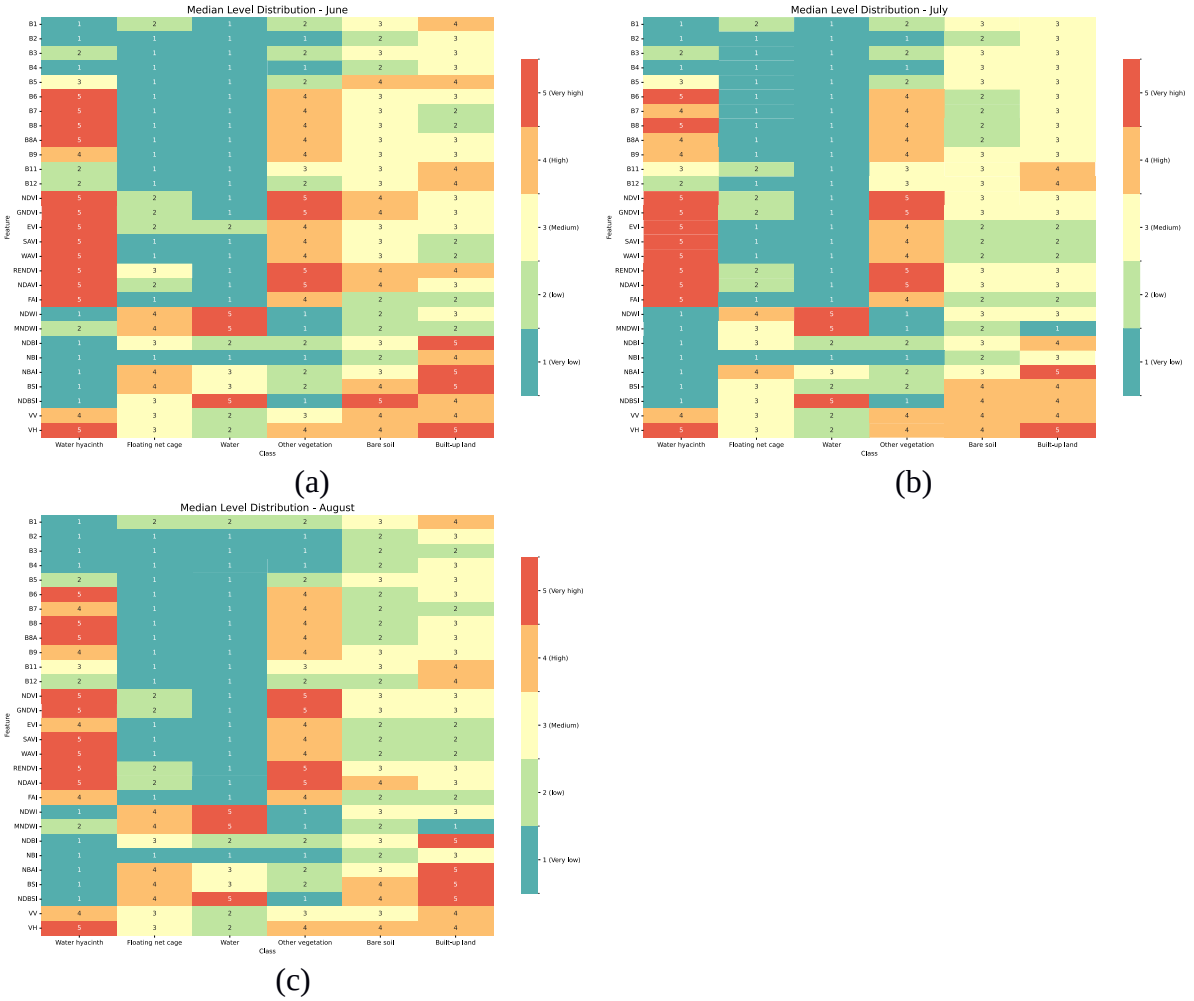
The 1D CNN was specifically chosen to utilize only spectral information, not spatial, ensuring a fairer methodological comparison with the Random Forest algorithm for the classification of six land cover classes. Both RF and a 1D CNN are pixel-based classifiers that operate on the spectral signature of an individual pixel without considering its neighbors. This approach allows us to isolate the effectiveness of the learning architecture itself—that is, comparing how a tree-based ensemble (RF) and a neural network (1D CNN) learn from the same vector of spectral information. A 2D CNN, by contrast, incorporates spatial context by analyzing patches of pixels, introducing additional information (texture, shape) that the RF model does not use. By focusing on a 1D CNN, this study first aims to establish the superiority of the CNN architecture in interpreting complex spectral patterns alone. (5) Evaluation, conducted to assess and compare the models' performance on the test data using cross-validation. The evaluation utilized several key metrics: overall accuracy, which measures the total percentage of correctly classified pixels; precision, which assesses the reliability of positive predictions; recall, which

measures the model's ability to detect all actual positive instances; and the F1-score, which provides a harmonic mean of precision and recall. The best model was then selected based on these metrics for the final stage; and (6) Deployment, in which the best-performing model was applied to the entire study area to generate a complete spatial classification. The results were then visualized as an interactive thematic map using an Earth Engine Web App, designed to effectively communicate the findings to stakeholders and support data-driven policy-making for the management of invasive water hyacinth in Rawa Pening Lake.

3. Result and Discussion

3.1. Feature identification

Descriptive analysis was used to identify the optimal features for land cover characterization from Sentinel-1 and Sentinel-2 satellite imagery and their derived spectral indices. All multispectral bands, radar bands, and spectral indices were normalized and standardized so that they had a range of values between 0 and 1 before descriptive analysis was performed. The next step is to divide each variable into 5 levels with Jenks Natural Breaks. The division of data into five levels aimed to group variables based on their value ranges, thereby revealing the distribution pattern for each variable. The Jenks Natural Breaks algorithm has the ability to divide data optimally by minimizing the variance of each class and on the other hand maximizing the variance between classes [22]. In validating the results of Jenks Natural Breaks, a Goodness of Variance Fit evaluation was used.



(a) (b)

(c)

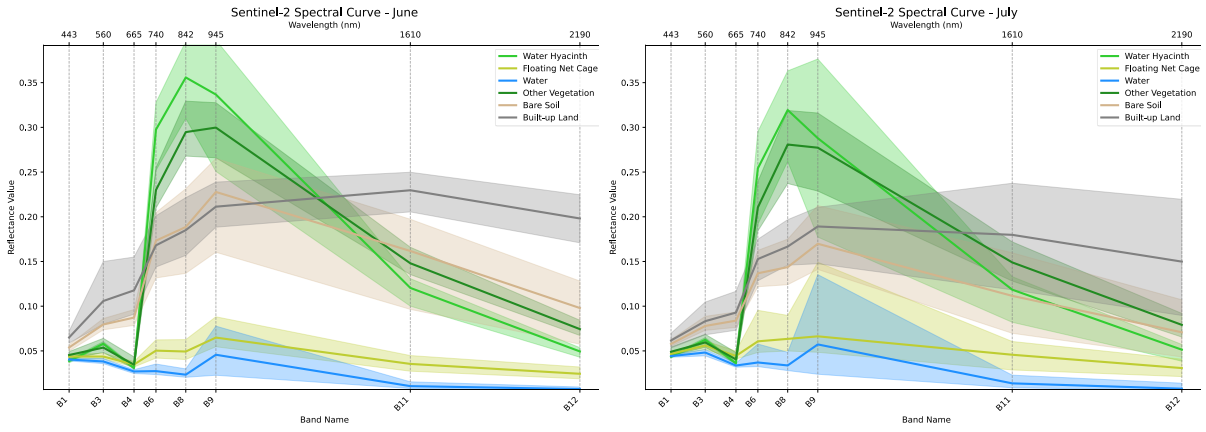
Figure 14. Heatmaps of median feature levels by land cover class for (a) June, (b) July, and (c) August.

Following the Jenks Natural Breaks classification, the heatmap of median feature levels (Figure 14) was used to perform a qualitative analysis of each variable's discriminatory power. The analysis focused on visually identifying 'characterizing variables' by assessing the uniqueness of their JNB level for each land cover class.

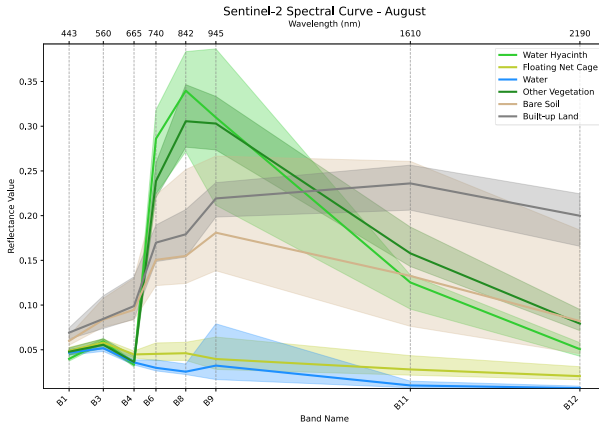
A variable was deemed a strong characterizer if its median JNB level for a particular class showed a clear separation from the others. The most powerful characterizers were those with an exclusive level. For example, by inspecting the June heatmap (Figure 14), several features are immediately identifiable as strong characterizers for Water Hyacinth, including B6, B7, B8, EVI, SAVI, and FAI, all of which register a 'Very High' (Level 5) value unique to that class. Conversely, indices like NDBI and BSI are shown to be strong characterizers for Water Hyacinth at the 'Very Low' level. This direct visual inspection of the heatmap provides a clear basis for understanding feature separability.

To confirm the characteristic variables, analysis was carried out using boxplots and spectral curves of each land cover, especially other vegetation and water hyacinths. The boxplot analysis is focused on finding out the difference in distribution between classes while the spectral curve analysis is focused on finding out the difference in the distribution between variables. Based on the analysis of the boxplot and spectral curves of each land cover, it can be seen that the distribution of values for several characteristic variables shows variations between months. Although in general the patterns in the same class are not much different.

The Sentinel-2 spectral curve (reflectance values at various wavelengths) provides a more detailed picture of the spectral response of each land cover in June, July, and August. On the Sentinel-2 spectral curve as seen in figure 7, the differences between the land cover classes are very pronounced at various wavelengths, and this pattern also shows temporal stability.



(a) (b)



(c)

Figure 15. Spectral reflectance curves of Sentinel-2 bands for different months: (a) June, (b) July, and (c) August.

The water class consistently displays very low reflectance values across the spectrum, with slight improvements towards blue wavelengths and strong absorption in Near-Infrared (B8) and Short-wave Infrared (B11 & B12). In contrast, water hyacinths and other vegetation exhibit a reflectance curve typical of vegetation, with strong absorption in the blue and red bands, as well as high peaks of reflectance in the NIR bands and decreases in SWIR. Specifically, water hyacinths exhibited higher peak reflectance in the NIR bands compared to other vegetation classes, confirming that water hyacinths have unique spectral characteristics that distinguish them from other vegetation. The curves for soils and built-up areas show relatively higher reflectance across the spectrum than for vegetation and water, with a gradual increase from the visible band to SWIR. Floating net cages display a spectral profile that sits between water and vegetation.

These spectral curve patterns were consistent in June, July, and August, both at the median and their distribution, suggesting that the spectral response of each land cover was a stable and robust feature for discrimination throughout the observation period.

Then to further deepen knowledge about the response and characteristics of the species of water hyacinth and other vegetation, a special boxplot analysis was carried out for the 2 classes. In general, the distribution of data levels for the observed classes (Water Hyacinth, KJA, Water, Vegetation, Soil, Built-up Areas) showed a strong pattern of similarity among the three images for June, July, and August. This consistency indicates that the main spectral characteristics of the various classes were fairly stable from month to month in this observation period.

The identification of the best features that characterize water hyacinths and other land cover is based on two complementary analysis approaches: Jenks Natural Breaks (JNB) level analysis, which highlights a variable's ability to uniquely isolate a class, and boxplot analysis and spectral curves that confirm the separability between classes even when the JNB levels overlap.

Based on a detailed analysis of JNB data, several variables were found that were consistently able to provide a unique level for a particular class. The MNDWI and NDWI Water Indices consistently place the Water class at level 5 (very high), while the other class is at a much lower level. This makes MNDWI, as a representative, a very reliable feature for separating water bodies. The EVI, SAVI, and FAI Vegetation Indices consistently place the water hyacinth class at level 5 (very high), a level that other classes do not have at most observation times. This confirms that the vegetation index is particularly strong in identifying the unique biomass density of water hyacinths. The NDBI and NBAI Building Indexes effectively isolate the Built-Up Areas class at level 5 (very high), proving its ability as a key characteristic.

Certain spectral bands also exhibit a unique ability to distinguish the cover class. The Near-Infrared (B8) band demonstrated a unique ability by placing the water hyacinth class at level 5, which was different from all the other classes in the June imagery. Similarly, SWIR bands such as B11 uniquely place water hyacinths at level 2 (low). On the other hand, there are crucial variables whose separability is more clearly visible through boxplot analysis, even though the JNB level is not entirely unique. The most significant example is in the Sentinel-1 radar band (VV and VH). JNB analysis showed that the VH level for water hyacinths (level 5) overlapped with the built-up areas (level 5). However, the boxplot analysis and the difference in JNB levels confirmed that the value of the VH backscatter in water hyacinths (level 5) was significantly higher compared to other vegetation (level 4) and KJA (level 3). A similar pattern is also seen in VV bands. This ability is essential to distinguish the physical structure of water hyacinths from other floating or terrestrial vegetation. Taking into account the results of these two analyses, the features selected for the model development stage are the most representative and have the highest separation, both uniquely and comparatively. MNDWI was chosen as a representation of the water index. FAI as a unique representation of vegetation index. NBAI as a representation of the build-up index. BSI as a representation of the soil index. Key spectral bands that exhibit the best separation ability: B8 (NIR) and B11 & B12 (SWIR) due to their ability to distinguish water hyacinth from other vegetation. VV and VH radar bands as vital structural markers to separate water hyacinths from other vegetation and KJA. The RGB (B4, B3, B2) and Red-Edge (B5) bands are retained for the completeness of basic spectral information.

3.2. Modeling

Two machine learning algorithms, namely Random Forest and Convolutional Neural Network (CNN), were used to build the best model for detecting water hyacinth cover. The model-building process for both algorithms involve searching for optimal hyperparameters using the Randomized Search technique and being evaluated through cross-validation to ensure good generalizations. The optimal hyperparameters selected for each model and month are detailed in Table 2 The results of cross-validation showed that the CNN model had a higher and more stable average accuracy (0.94 for June, 0.91 for July, and 0.95 for August) compared to Random Forest (0.94 for June, 0.87 for July, and 0.92 for August).

Table 2. Hyperparameter model.

CNN			
Hyperparameter	**Value (June)**	**Value (July)**	**Value (August)**
Convolutional filter	8	128	128
Convolutional kernel size	5	5	3
Dense unit	128	64	64
Batch size	8	8	32
Epochs	50	50	50
Kernel initializer	He normal	Lecun normal	Glorot uniform

Random Forest			
Hyperparameter	**Value (June)**	**Value (July)**	**Value (August)**
N tree	50	50	50
Max depth	15	13	17
Min samples split	4	4	4
Min samples leaf	1	2	2

Based on performance evaluations of test data using F1 accuracy, precision, recall and score metrics, the CNN model consistently shows better performance. The detailed results of this evaluation are presented in Table 3.

Table 3. Results of machine learning model evaluation.

Algorithm	Month	Accuracy	Precision	*Recall*	F1 Score
Random Forest	June	0.96	0.96	0.96	0.96
CNN		0.96	0.96	0.96	0.96
Random Forest	July	0.89	0.89	0.89	0.89
CNN		0.92	0.92	0.92	0.92
Random Forest	August	0.94	0.94	0.94	0.94
CNN		0.97	0.97	0.97	0.97

In the June image data, both models had identical performance with a score of 0.96 in all metrics. However, in July and August, the CNN model demonstrated markedly better performance. For July, CNN achieved a score of 0.92, while Random Forest scored 0.89. In August, CNN's lead became clearer with a score of 0.97 compared to Random Forest's 0.94. This consistent outperformance across multiple metrics and months supports its selection as the more robust model. The CNN's superiority in this task can be attributed to its internal mechanism of hierarchical feature learning. Unlike Random Forest, which treats each spectral band and index as an independent variable, a 1D CNN treats the spectral signature of a pixel as a sequence. The convolutional layers act as learnable filters that slide across this spectral sequence, automatically identifying complex local patterns such as the steep slope of the red edge or specific absorption features that are characteristic of different land cover types. Deeper layers in the network can then combine these simple patterns into more abstract and powerful features. This ability to automatically learn relevant features from the raw spectral data gives the CNN an advantage over the tree-based approach of Random Forest, leading to its better performance in this classification task.

3.3. Mapping

Mapping of water hyacinth cover was carried out using the best model, namely CNN. Based on the results of the classification of satellite imagery for the months of June, July, and August, the number of pixels of water hyacinth cover was obtained in each month. The number of pixels detected by water hyacinths in each month is as follows:

June : 34,775 pixels

July : 32,627 pixels

August : 34,175 pixels.

In addition, the mapping also includes pixel calculations for other land cover classes in the study area. Table 4 shows the pixel aggregation results for each land cover class identified in the study area.

Table 4. Number of pixels of land cover per class in the research study area

Class	Pixel count		
	June	July	August
Water hyacinth	34,775	32,627	34,175
Floating net cage	15,095	7,505	12,563
Water	103,667	111,016	103,863
Other Vegetation	2,921	4,772	4,852
Soil	1,319	1,071	1,935
Built-up areas	500	1,286	889

After obtaining the results of the detection of water hyacinth cover pixels and other land cover classes, the next step is to visualize the classification results using the best model. This mapping aims to provide a clear spatial picture of the distribution of water hyacinths and other land cover in the study area. The results of this mapping were obtained by applying the best models that have been trained in advance on satellite imagery. Each class of land cover, including water hyacinths, is represented by a different color on the map for easy visual identification. Figure 16 shows the mapping results of classified land cover, including the distribution of water hyacinths.

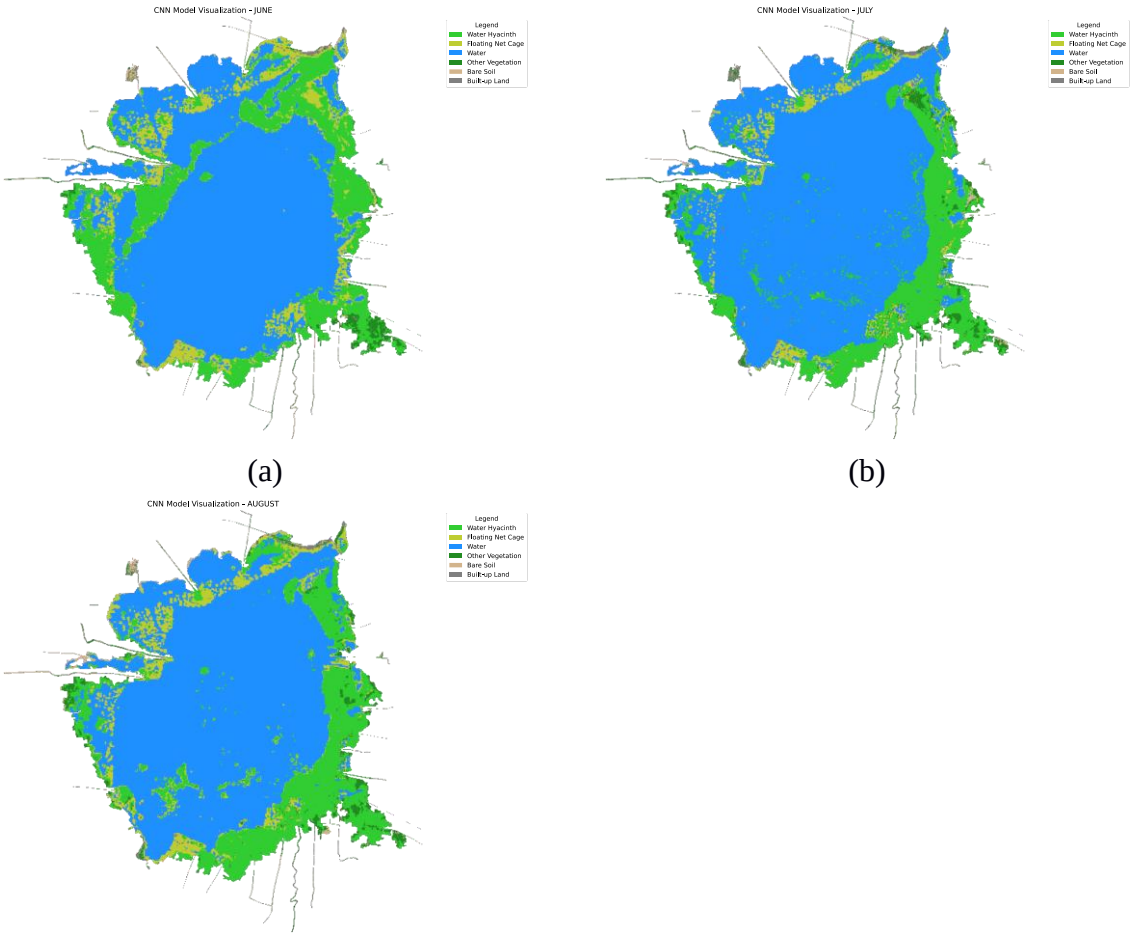


Figure 16. Land cover distribution maps for (a) June, (b) July, and (c) August.

On the resulting map as shown in Figure 16, each pixel classified as a water hyacinth will be marked in light green, providing a clear picture of its location and distribution in the study area. Thus, this map provides visual information that is very useful in understanding the distribution of water hyacinths and other land cover classes in the region.

The fluctuation in water hyacinth pixel counts between June (34,775), July (32,627), and August (34,175) suggests dynamic changes in the lake's ecosystem. The notable decrease in July corresponds with documented manual cleaning efforts by local authorities, which were confirmed by the available aerial photography for that month. The subsequent increase in August indicates the rapid regrowth potential of water hyacinth, highlighting the need for continuous monitoring.

The classification maps (Figure 16) reveal that water hyacinth is not uniformly distributed but tends to concentrate along the lake's shallower edges and in calmer bays, particularly in the northern and southern parts of the lake. This pattern was consistent across all three months, although the density in specific areas varied, likely influenced by wind, currents, and management interventions.

These findings have significant practical implications. The high-resolution monthly maps can serve as a critical tool for the sustainable management of Rawa Pening Lake. Lake authorities, such as BBWS Pemali Juana, can use this data to deploy cleaning equipment and resources more efficiently to target areas with the highest concentration of water hyacinth, moving from a reactive to a proactive management strategy.

4. Conclusion

This research established a robust, high-accuracy methodology for monitoring invasive water hyacinth, providing a critical tool for the rapid and effective surveillance necessary to anticipate and mitigate large-scale blooming events. The study successfully addressed its core objectives by first identifying a unique combination of vegetation indices and radar bands, such as FAI and VH polarization, that effectively discriminate water hyacinth from other complex land cover classes, including different vegetation types and aquaculture structures. Following this, the research validated the superiority of a one-dimensional Convolutional Neural Network (CNN) model, which consistently outperformed the Random Forest algorithm in classification accuracy. The application of this validated CNN model produced precise water hyacinth distribution maps, demonstrating its capability to effectively track the monthly fluctuations of water hyacinth distribution. The implications of this work are significant, offering a scalable and reliable remote sensing framework to support the sustainable management of Indonesia's priority inland water ecosystems.

However, a key limitation of this study is the verification method for the ground truth data. While the July classification is strongly supported by high-resolution aerial photography, the accuracy of the June and August maps relies on visual interpretation of satellite imagery. Therefore, the findings for these months should be considered preliminary.

To build upon these findings, future research should focus on enhancing both data integrity and model architecture. A foundational improvement would be to incorporate rigorous in-situ ground truth verification to ensure the training data accurately represents real-world conditions, while also extending the analytical timeframe to a full annual cycle to capture seasonal patterns and more precisely predict peak blooming periods. In terms of model development, now that this study has demonstrated the superior ability of the CNN architecture to learn from spectral data at the pixel level, the logical next step is to implement a two-dimensional CNN (2D CNN). Such a model would leverage the spatial context that was intentionally excluded from this study's comparison. By analyzing the texture and shape of water hyacinth mats from neighboring pixels, a 2D CNN could further improve classification accuracy. Concurrently, future work should explore more computationally efficient models that utilize only the most significant features identified in this study, such as the FAI index and VH radar band, to optimize the methodology for broader, large-scale applications.

Acknowledgements

The authors would like to express their sincere gratitude to all parties who supported the completion of this research.

Special appreciation is conveyed to the Balai Besar Wilayah Sungai (BBWS) Pemali Juana for providing the aerial photography data of Lake Rawa Pening. This contribution was indispensable for the data labeling process and the overall success of this study.

Gratitude is also extended to the academic community of the D-IV Statistical Computing Study Program at Politeknik Statistika STIS for providing the facilities and educational environment that made this work possible. Heartfelt thanks are also due to colleagues and friends for their companionship, constructive discussions, and encouragement.

Finally, the authors wish to extend their deepest gratitude to their beloved families for their endless love, prayers, and support.

References

[1] L. Pádua *et al.*, "Water hyacinth (Eichhornia crassipes) detection using coarse and high resolution multispectral data," *Drones*, vol. 6, no. 2, p. 47, Feb. 2022, doi: 10.3390/drones6020047.

[2] R. Gerardo and I. P. de Lima, "Assessing the potential of Sentinel-2 data for tracking invasive water hyacinth in a river branch," *J Appl Remote Sens*, vol. 16, no. 1, Feb. 2022, doi: 10.1117/1.JRS.16.014511.

[3] G. Singh, C. Reynolds, M. Byrne, and B. Rosman, "A remote sensing method to monitor water, aquatic vegetation, and invasive water hyacinth at national extents," *Remote Sens (Basel)*, vol. 12, no. 24, p. 4021, Dec. 2020, doi: 10.3390/rs12244021.

[4] Department of Economic and Social Affairs, "Goal 15 | Department of Economic and Social Affairs," United Nations. Accessed: Dec. 30, 2024. [Online]. Available: https://sdgs.un.org/goals/goal15#progress_and_info

[5] Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia, *Peraturan Menteri Lingkungan Hidup dan Kehutanan Nomor P.94/MENLHK/SETJEN/KUM.1/12/2016 Tahun 2016 tentang Jenis Invasif*. 2016. Accessed: Dec. 30, 2024. [Online]. Available: https://ksdae.menlhk.go.id/assets/news/peraturan/PERMEN_LHK_N0_P._94_TH_2016_TTG_JENIS_INVASIF_.pdf

[6] J. Rojas-Sandoval and P. Acevedo-Rodríguez, "Eichhornia crassipes (water hyacinth)," Jul. 07, 2013. doi: 10.1079/cabicompendium.20544.

[7] L. Navarro and G. Phiri, *Water hyacinth in Africa and the Middle East : A survey of problems and solutions*. International Development Research Centre, 2000. Accessed: Jul. 28, 2025. [Online]. Available: https://publications.gc.ca/site/eng/9.696841/publication.html

[8] Presiden Republik Indonesia, *Peraturan Presiden Republik Indonesia Nomor 60 Tahun 2021 tentang Penyelamatan Danau Prioritas Nasional*. 2021. Accessed: Dec. 30, 2024. [Online]. Available: https://jdih.maritim.go.id/cfind/source/files/perpres/2021/salinan-perpres-nomor-60-tahun-2021-new.pdf

[9] A. W. Utomo, "Merajut hidup dari bengok: Pola-pola pemanfaatan bengok (eceng gondok) di sekitar Danau Rawa Pening dalam perspektif pembangunan berkelanjutan," *Cakrawala Jurnal Penelitian Sosial*, vol. 5, no. 2, pp. 191–216, Jan. 2017, Accessed: Dec. 30, 2024. [Online]. Available: https://ejournal.uksw.edu/cakrawala/article/view/666

[10] P. W. Purnomo, P. Soedarsono, and M. N. Putri, "Profil vertikal bahan organik dasar perairan dengan latar belakang pemanfaatan berbeda di Rawa Pening," *Management of Aquatic Resources Journal (MAQUARES)*, vol. 2, no. 3, pp. 27–36, Aug. 2013, doi: 10.14710/marj.v2i3.4178.

[11] Kementerian Lingkungan Hidup dan Kehutanan, "Profil Danau Rawa Pening." Accessed: Dec. 14, 2024. [Online]. Available: https://www.profildanau.info/danau/rawapening

[12] S. B. Bricker, C. G. Clement, D. E. Pirhalla, S. P. Orlando, and D. R. G. Farrow, "National estuarine eutrophication assessment: Effects of nutrient enrichment in the nation's estuaries," 1999. Accessed: Dec. 14, 2024. [Online]. Available: https://repository.library.noaa.gov/view/noaa/1693

[13] Badan Pusat Statistik Kabupaten Semarang, "Jumlah pengunjung di tempat rekreasi wisata buatan Kabupaten Semarang." Accessed: Jul. 02, 2025. [Online]. Available: https://semarangkab.bps.go.id/id/statistics-table/2/ODE4IzI=/jumlah-pengunjung-di-tempat-rekreasi-wisata-buatan-kabupaten-semarang--jiwa-.html

[14] T. Dube, O. Mutanga, M. Sibanda, V. Bangamwabo, and C. Shoko, "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*, vol. 84, pp. 11–22, Jul. 2017, doi: 10.1016/j.apgeog.2017.04.005.

[15] I. K. Felix, M. D. Simpson, and A. Marino, "Detecting and mapping of water hyacinth in Lake Victoria using radar polarimetric data," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Jul. 2023, pp. 7210–7213. doi: 10.1109/IGARSS52108.2023.10282297.

[16] A. Datta *et al.*, "Monitoring the spread of water hyacinth (Pontederia crassipes): Challenges and future developments," *Front Ecol Evol*, vol. 9, Jan. 2021, doi: 10.3389/fevo.2021.631338.

[17] E. C. Rodríguez-Garlito, A. Paz-Gallardo, and A. Plaza, "Monitoring the spatiotemporal distribution of invasive aquatic plants in the Guadiana River, Spain," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 16, pp. 228–241, 2023, doi: 10.1109/JSTARS.2022.3225201.

[18] E. C. Rodríguez-Garlito, A. Paz-Gallardo, and A. Plaza, "Automatic detection of aquatic weeds: A case study in the Guadiana River, Spain," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 15, pp. 8567–8585, 2022, doi: 10.1109/JSTARS.2022.3210373.

[19] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Practical Application of Knowledge Discovery and Data Mining*, Practical Application Co, Apr. 2000, pp. 29–40.

[20] P. Mata, "Classifying land cover of Ireland from satellite imagery using deep learning and transfer learning," Master's thesis, National College of Ireland, Dublin, 2023.

[21] C. Haub, L. Kleinewillinghöfer, V. Garcia Millan, and A. Di Gregorio, "Protocol for land cover validation," Apr. 2015.

[22] B. D. Dent, J. S. Torguson, and T. W. Hodler, *Cartography: Thematic map design*, 6th ed. McGraw-Hill Higher Education, 2009.