Detection and Mapping of Water Hyacinth Growth Cycle in Anzali International Wetland Using Sentinel-2 Time Series

Mehran Alizadeh Pirbasti, *Student Member, IEEE*, Vahid Akbari, *Senior Member, IEEE*, Deepayan Bhowmik, Savitri Maharaj, and Armando Marino, *Member, IEEE*

***Abstract*—Water hyacinth (WH) is a notorious invasive species that significantly threatens ecosystems worldwide. Despite WH's well-documented threats and effects, its spatial distribution is not yet fully understood, especially in complex environments such as wetland systems. This knowledge gap is primarily due to the lack of accurate techniques with high spatial resolution and reliable in situ field data for quantification and monitoring. To address this research gap, we conducted a study to map the spatiotemporal distribution of invasive WH in Anzali International Wetland, Iran, using Sentinel-2 Multispectral Instrument 2022 data. Specifically, our study aimed to identify multispectral remote sensing variables and in situ field data using machine learning (ML) methods to detect and map WH growth cycles. In the first phase of our study, we compared three ML models for detecting WH and discriminating from other classes. Our results demonstrate that ML algorithms can detect WH accurately. In the second phase, we used four images dominated by four growth stages: early, mid, high, and decaying stages to train our ML classifier. We used the random forest algorithm for training our training samples achieving an overall classification accuracy of over 98%. These findings were further supported by statistical analysis, such as F1 (above 96%) and intersection over union (above 92%), indicating the high-performance quality of the used algorithm. Our study provides valuable insights into using ML algorithms for mapping WH growth cycles, which can significantly contribute to effectively managing and monitoring invasive species worldwide.**

***Index Terms*—Anzali wetland, growth cycle, machine learning (ML), random forest (RF), Sentinel-2 (S2) time series, water hyacinth (WH).**

Nomenclature

Abbreviation	*Definition*
AIW	Anzali International Wetland.
S2	Sentinel-2.
NDVI	Normalized Difference Vegetation Index.
DA	Discriminant analysis.
ML	Machine learning.
RF	Random forest.
MLP	Multilayer perceptron.
SGD	Stochastic gradient descent.
CNN	Convolutional neural network.
OV	Other aquatic vegetation.
SW	Shallow water.
DW	Deep water.
TV	Terrestrial vegetation.
WH	Water hyacinth.
WH1	WH in early stage.
WH2	WH in mid stage.
WH3	WH in high stage.
WH4	WH in decaying stage.
OA	Overall classification accuracy.
AUC	Area under the curve.
IoU	Intersection over union.

Manuscript received 26 September 2023; revised 11 April 2024; accepted 26 June 2024. Date of publication 11 July 2024; date of current version 5 August 2024. This work was supported in part by the Royal Academy of Engineering through the Frontiers Follow-on Funding Scheme under Grant (FF\1920\1\37). *(Corresponding author: Mehran Alizadeh Pirbasti.)*

Mehran Alizadeh Pirbasti is with the SFI Centre for Research Training in Machine Learning at School of Computer Science, University College Dublin D04 V1W8 Dublin, Ireland (e-mail: mehran8official@gmail.com).

Vahid Akbari and Savitri Maharaj are with the School of Computing Science and Mathematics, University of Stirling, FK9 4LA Stirling, U.K.

Deepayan Bhowmik is with the School of Computing, Newcastle University, Newcastle Upon Tyne NE4 5TG, U.K.

Armando Marino is with the Department of Environmental and Biological Sciences, The University of Stirling, FK9 4LA Stirling, U.K.

Digital Object Identifier 10.1109/JSTARS.2024.3427002

I. Introduction

INVASIVE species can threaten biodiversity, ecological stability, and human well-being in natural ecosystems worldwide [1], [2]. Among them, *Eichhornia crassipes*, commonly known as WH, is one of the most notorious aquatic invaders due to its rapid growth, high reproductive capacity, and ability to form dense mats that impede water flow and reduces dissolved oxygen levels [3].

WH can cause significant ecological and economic damage to wetland ecosystems [4]. WH has been reported as a significant problem in Iran in AIW, the largest and most important wetland in the Caspian Sea region [5]. The AIW is a critical Ramsar site and a major ecological hot spot in the Caspian Sea basin, hosting a rich diversity of flora and fauna and providing various ecosystem services to local communities [6]. However, the invasion of WH has become a severe challenge to the conservation and management of the AIW, threatening its biodiversity, water quality, fishery, tourism, and other benefits [7]. Due to WH's rapid growth and spread, monitoring and mapping the growth stages of WH in the wetland is crucial for developing effective management strategies to mitigate its negative impacts.

Manual surveys involve visually inspecting water bodies to identify the presence and distribution of WH and other water plants. These surveys are typically conducted by trained personnel who travel along the shoreline or in boats, visually inspecting the water surface for signs of invasive species. The surveys can be conducted regularly to monitor changes in the distribution and density of water plants over time. Traditional methods for monitoring WH, such as field surveys and aerial photography, are time consuming, labor intensive, and expensive and may not provide up-to-date or comprehensive information [8].

In recent years, remote sensing has emerged as a powerful tool for monitoring invasive species in aquatic environments, providing synoptic and repetitive coverage of large areas with high spatial and spectral resolution [9]. To effectively manage and control the spread of WH in AIW, accurate and timely information on its distribution, growth, and dynamics is essential [10]. Therefore, there is a growing interest in using remote sensing data and ML techniques to monitor invasive species in wetlands, as they offer advantages in cost-effectiveness, spatial and temporal coverage, and accuracy [11].

Remote sensing satellites can monitor WH and other water plants through various sensors, including multispectral, hyperspectral, and radar sensors. Multispectral sensors capture data in multiple wavelengths of the electromagnetic spectrum, allowing for identifying different types of vegetation based on their spectral signatures [12]. Hyperspectral sensors capture data in a more significant number of narrow spectral bands, providing greater detail on vegetation composition [13], [14]. Radar sensors can penetrate vegetation canopies to capture information on plant structure and biomass [15], [16].

Recently, ML algorithms have shown great potential for accurately detecting and mapping WH in wetland ecosystems [17]. ML algorithms have been increasingly used for remote sensing applications, including detecting and mapping vegetation cover in wetland ecosystems [18], [19]. ML algorithms can be trained using data from remote sensing satellites, which capture plant distribution and density data. These data can be used to identify patterns and relationships that can be used to accurately predict the presence and growth of WH and other water plants. Various ML techniques can be used for this purpose, including decision trees, RFs, and neural networks [20].

Several studies have demonstrated the effectiveness of remote sensing and ML for WH detection and mapping of different wetlands around the world, including Lake Victoria in East Africa [21], the Río de la Plata Turbid in Argentina [22], and Lake Tana in Ethiopia [18]. Previous studies have demonstrated the effectiveness of ML-based approaches for detecting WH in other water bodies [23], [24].

Sibanda et al. [25] used the DA to classify WH in two dry and wet seasons from Landsat 8 images. They achieved an overall accuracy of 95% and highlighted the effectiveness of or accurately detecting WH growth. In another study, Mukarugwiro et al. [26] used an RF algorithm to map spatiotemporal changes of WH in Rwandan water. They used time series of Landsat images and achieved an overall accuracy of 87%. The study highlighted the potential of RF for WH detection and provided valuable information for effectively managing the weed.

Rodríguez-Garlito et al. [27] used a CNN model to detect WH by S2; they found that a comprehensive training set of WH at different stages of its biological life cycle could be collected by visually interpreting the S2 images. They also demonstrated that the CNN model only needed to be trained once using a high-resolution image to extract the most representative samples in the study area, allowing invasive plants to be separated from other land cover surfaces. Also, in another work, they developed a new CNN method for detecting invasive aquatic plants in S2 images using different spectral indices. The researchers found that training the CNN model with spectral indices provided better detection accuracy than training it with all available spectral bands. The article suggests that this new method could lead to more efficient and effective data processing [28].

No single ML algorithm can handle all pattern recognition tasks due to differences in texture and spectral reflectance of the ground features, and different algorithms often produce variable results [29]. Moreover, the relative performance of various classification methods will vary by season due to changes in the physiological characteristics of plants and the life cycle influence on spectral reflectance. However, accuracy is crucial for evaluating the performance of a classifier. Although WH threats and effects have been thoroughly documented, its spatial distribution is still poorly understood. Researchers continue to strive to identify WH, as previous research in this field has yielded insufficient outcomes. This persistent pursuit is driven by several compelling reasons, which are as follows.

1) *Lack of in situ field data:* One significant challenge in mapping WH growth cycles is the lack of reliable in situ field data. In many cases, inaccessible areas or insufficient resources for access make it difficult to collect in situ field data. Moreover, due to the rapid growth of WH, continuous monitoring is required, which makes in situ field data collection not cost-effective. The absence of in situ field data can lead to inaccurate classification results, which can affect the overall performance and reliability of the algorithm.
2) *Cloudy images:* The presence of clouds in optical data is another challenge in mapping the WH. Clouds can obstruct the view of the earth's surface, making it difficult for satellites to accurately measure changes in WH that are influenced by seasonal cycles. Cloud cover can also vary spatially and temporally, leading to incomplete datasets that can affect the accuracy of the classification results.

We implemented a systematic field observation approach to overcome the limitations posed by cloud coverage in our study area. This involved conducting weekly field visits to the area of interest, which allowed us to collect supplementary data to compensate for the lack of clear satellite images. The primary objectives of these field observations were as follows:

1) to obtain in situ field data for the validation of remote sensing results;
2) to identify potential discrepancies between satellite images and land surface conditions;
3) to ensure a continuous flow of information to support our analysis, regardless of the availability of cloud-free satellite images.

During each field visit, we collected data on vegetation cover, land use, and other relevant environmental parameters using a combination of direct measurements, photographic documentation, and field notes. These field observations were then used to augment the remote sensing data, allowing us to understand the study area comprehensively despite the challenges posed by cloud coverage.

Therefore, our primary objective is to prepare a spatial–temporal distribution map of WH in AIW using the ML classifier. The secondary goal is the identification of multispectral remote sensing variables and in situ field data using ML to detect and predict the growth stages of WH. While previous studies have examined the use of remote sensing and other technologies to monitor vegetation growth cycles, none have specifically focused on using ML algorithms to map the growth cycles of WH. Therefore, our study represents a novel approach that can contribute to advancing scientific knowledge in this area and has practical implications for management and planning.

This work contributes in two significant ways. First, we have developed a novel approach that combines ML algorithms and S2 satellite data to accurately detect and map the spatiotemporal distribution of WH in a complex wetland environment. Our approach offers a reliable and efficient solution for monitoring WH, providing valuable insights for decision-making processes and effective management strategies. Second, our research has uncovered exciting findings regarding the spectral signature of WH during different stages of its growth stage and its NDVI. We demonstrated that the spectral signature of WH in each growth stage has distinct differences, enabling the detection of different growth cycles using a single image. We demonstrated the effectiveness of the RF algorithm in detecting WH growth stages with high accuracy. This finding highlights the potential of ML techniques in accurately identifying and tracking the growth stages of WH, enabling timely interventions and management actions. Our study contributes to environmental monitoring and management by providing a practical and reliable approach for mapping WH spatial distribution and growth stages. The insights gained from this research can support decision makers and conservationists in implementing effective strategies to mitigate the impact of invasive species on ecosystems.

The rest of this article is organized as follows. Section II describes the study area, in situ field and satellite datasets, data labeling, and methodology. Results and discussions are presented in Section III. Finally, Section IV concludes this article.

II. Materials and Methods

A. Study Area

AIW (see Fig. 1), located in northern Iran, is one of the most significant wetlands in the Middle East and a designated UNESCO Biosphere Reserve. It covers an area of approximately 19 000 hectares and is a critical habitat for many endangered species, including birds, fish, and reptiles. However, it is currently included in the Montreal List because the wetland faces significant threats, including pollution, habitat loss, and the invasive WH.

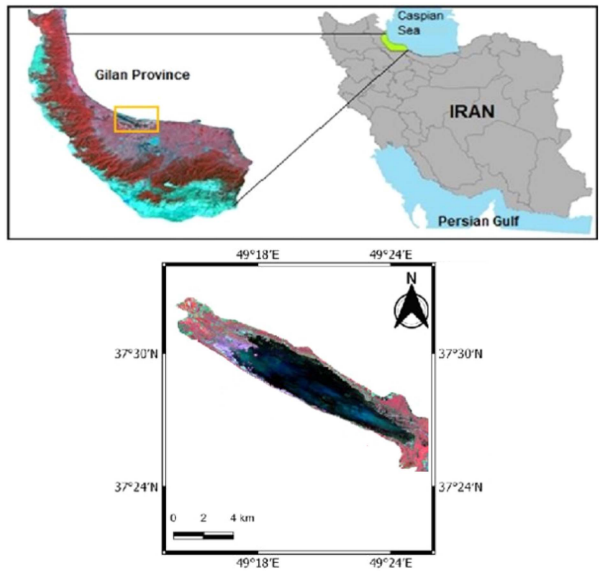

Fig. 1. Top: Our study area located in Guilan Province, Iran. Bottom: False color composite of S2 imagery over AIW.

Fig. 2. Field observations of WH growth stages. From left to right: early stage, mid stage, high stage, and decaying stage.

WH was first introduced to the AIW in 2013 and has since spread rapidly, leading to significant ecological and economic impacts. The plant forms dense mats on the water surface, reducing light penetration and oxygen exchange, which can result in the death of aquatic fauna and flora. The presence of WH also impacts local fishing and agriculture industries, hindering water transportation and irrigation activities. A study found that parts of AIW with a stable presence of WH daily had a 0.2-cm drop in water height, 2.5 times more than the areas without WH cover [30].

B. In Situ Field Data and Satellite Datasets

To accurately map the spatial distribution of WH in the AIW, we conducted regular visits to identify sites infested with WH and observe different growth stages simultaneously. During these visits, we collected data on different attributes of WH, including its flowering condition, growth stage, vertical stem, and corresponding S2 data with cloud percentage. Table I presents the different attributes of WH and corresponding S2 data with minimal cloud cover used for mapping WH growth stages. Fig. 2

TABLE I
FIELD OBSERVATIONS OF WH GROWTH STAGES IN AIW WITH CLOUD PERCENTAGE OF S2 IMAGES AND CHOSEN S2 IMAGES WITH MINIMAL CLOUD COVER FOR MAPPING WH GROWTH CYCLE

Growth stage	WH height (cm)	Acquisition time	Flowering	Cloud percentage	Used S2
Low	4	17/03/2022	No flowering	99.9	
Low	5	24/03/2022	No flowering	100.0	
Low	7.5	31/03/2022	No flowering	20.1	31 March 2022
Low	8	7/4/2022	No flowering	54.3	
Mid	10	14/04/2022	No flowering	70.5	
Mid	12	21/04/2022	No flowering	78.2	
Mid	14	28/04/2022	No flowering	38.7	25 April 2022
Mid	17	5/5/2022	No flowering	71.3	
Mid	20	12/5/2022	Flowering start	96.0	
Mid	23	19/05/2022	Flowering	9.78	15 May 2022
Mid	27	26/05/2022	Flowering	94.9	
High	39	2/6/2022	Flowering	0.0	04 June 2022
High	47	9/6/2022	Flowering	24.2	
High	54	16/6/2022	Flowering	0.0	14 June 2022
High	62	23/6/2022	Flowering	20.6	
High	70	30/6/2022	Flowering finish	57.7	
High	78	7/7/2022	No flowering	12.8	
High	78	14/7/2022	No flowering	0.0	14 July 2022
High	78	21/7/2022	No flowering	92.0	
Decaying	110	25/11/2022	No flowering	28.1	26 November 2022
Decaying	120	20/12/2022	Plant death	93.3	

illustrates the progression of WH growth stages at various time points.

C. Data Labeling

To accurately map the spatial distribution of WH in AIW, we identified different classes inside and outside the wetland. The classes identified were WH, OV, DW, SW, and TV. The classification system we used aimed to discriminate WH from other classes. To achieve this, we used labeling data to train the classification system. The data labeling stage consisted of three parts.

First, a hand-held GPS was utilized to collect ground coordinates from diverse classes, resulting in a total of 182 geographical coordinates. These coordinates served as reference points for the different classes present in the study area.

Second, as shown in Fig. 3, manual identification and labeling are conducted. By manually labeling the point samples, the classification system could learn the distinguishing characteristics of each class. In total, 8250 labels were assigned, covering areas with WH, OV, DW, SW, and TV. The manual labeling process ensured that the classification system learned the distinguishing characteristics of each class. Finally, the labeled areas were cross-checked with the ground coordinates obtained in the first stage. This step ensured that the classification system was trained using precise and reliable data, aligning the labeled areas with the corresponding in situ field data coordinates.

This approach of point sampling and manual labeling is a reasonable way to label the data for several reasons. First, in situ field data collection through GPS observations allows us to gather information from specific locations within the study area. This approach ensures that the labeled data represent the actual classes present in the wetland environment, improving the accuracy and reliability of the classification system.

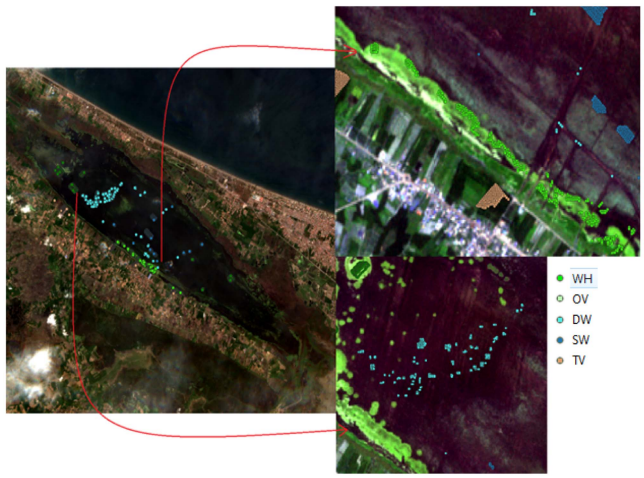

Fig. 3. Example of point sampling inside AIW, where WH = water hyacinth, OV = other aquatic vegetation, SW= shallow water, DW = deep water, and TV= terrestrial vegetation.

Second, manual identification and labeling enable us to visually discern and label different areas based on their characteristics. This approach uses human expertise to distinguish between classes, especially when dealing with complex and heterogeneous environments like wetlands. Manual labeling allows us to capture the subtle variations and patterns in the imagery that other methods might miss, such as changes in the plant growth cycle, water depth, and other characteristics, resulting in more accurate classification results. By employing this data labeling process, we ensured that the classification system was trained using precise and reliable data, enhancing its ability to classify the various classes in the data accurately.

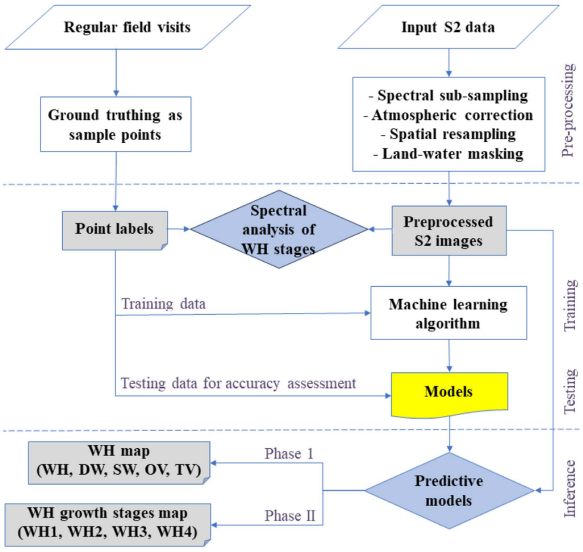

Fig. 4. Flowchart of our method used for WH mapping, where WH1 = WH in early stage; WH2 = WH in mid stage; WH3 = WH in high stage; and WH4 = WH in decaying stage.

TABLE II
CLASSIFICATION RESULTS OF MLP, SGD, AND RF CLASSIFIERS

MLP

Class	F1	Recall	Precision	IoU
WH	90.3	89.0	91.6	82.3
OV	94.0	94.4	93.7	88.7
SW	90.0	94.8	85.8	81.9
DW	73.9	66.7	82.9	58.6
TV	99.1	99.1	99.1	98.2
Overall accuracy: 87.5%, Kappa coefficient: 0.867				

SGD

Class	F1	Recall	Precision	IoU
WH	88.9	91.3	86.7	80.1
OV	90.9	89.6	92.3	83.4
SW	90.3	94.4	86.6	82.4
DW	71.7	65.5	79.2	55.9
TV	94.0	91.0	97.2	88.7
Overall accuracy: 89.2%, Kappa coefficient: 0.895				

RF

Class	F1	Recall	Precision	IoU
WH	90.5	90.5	89.0	91.0
OV	94.7	94.0	95.4	90.0
SW	94.7	98.4	91.3	90.0
DW	88.5	83.9	93.6	79.3
TV	97.4	95.9	99.0	95.0
Overall accuracy: 91.5%, Kappa coefficient: 0.918				

D. Methodology

Fig. 4 graphically shows the workflow adopted in this research. The process begins with acquiring multispectral images over the AIW via the S2 satellite. We procured S2's Multispectral Instrument Level-2A data from the Copernicus Open Access Hub.

Following the data acquisition, several preprocessing techniques are applied to the images. These techniques include spectral subsampling, atmospheric correction, spatial resampling, and combining of the 10-m bands from the S2 data, comprising the red (R), green (G), blue (B), and near-infrared (NIR) bands, and masking land-water area.

Subsequently, the preprocessed images are then extracted for the case study area. The data are then labeled according to the protocol outlined in Fig. 3, which involves the manual identification and point labeling of the various classes available for the training and validation of the classification system.

In this research, we employ three classification algorithms, i.e., RF, SGD, and MLP, chosen for their demonstrated efficacy in past studies [31], [32], [33]. The algorithm that yields the highest performance is selected for the secondary objective.

We then proceed to analyze the accuracy of the classification results using diverse evaluation metrics, such as overall accuracy, the Kappa coefficient, and F1-score, as well as recall, precision, IoU, and the receiver operating characteristic (ROC) curve. Subsequently, ML models are utilized to generate an automated map of WH based on the classification results.

Simultaneously, we collect and analyze the spectral signatures and NDVI at various growth stages as part of the secondary objective. This analysis provides valuable insights into the spectral characteristics and vegetation dynamics associated with WH, which can be used for inference. In the final stage, an automated map of WH growth stages is generated using the chosen ML model. We evaluate the outputs by comparing them with the in situ field data, using calculating different accuracy measures. The methodology employed in this study has been designed to be robust and adaptable under varying environmental conditions and data availability.

III. RESULTS AND DISCUSSION

A. Phase I: WH Detection

The classification task in this phase was performed on a single image containing five classes: WH, DW, SW, TV, and OV. The workflow of the study is presented in Fig. 4. The dataset used for this classification comprised 8250 samples, with 6600 samples allocated for training the ML algorithms and 1650 samples for testing and validation. This dataset division allowed for robust evaluation of the classification algorithms' performance on unseen samples. Three ML algorithms were evaluated, namely, RF, MLP, and SGD, which were found to perform the best in classification outcome. The RF algorithm was configured with 700 estimators and a depth of 5, while the MLP algorithm was configured with six layers of neurons and a batch size of 8. The SGD algorithm was configured with a learning rate of alpha = 0.02; to enhance the performance of the MLP and SGD algorithms, we applied min–max normalization to the input data.

Based on their classification accuracy, the algorithms were compared. The RF algorithm was found to be the most accurate in detecting WH, achieving an overall accuracy of 91.5%. The detailed results of the three algorithms evaluations and accuracies of the different classes are presented in Table II. The confusion matrix for the three algorithms is shown in Fig. 5, which indicates the number of correctly and incorrectly classified pixels for each class. The ROC curves for the three algorithms are shown in Fig. 6, which demonstrate the tradeoff between sensitivity and specificity for different classification thresholds. The output models for the three algorithms are shown

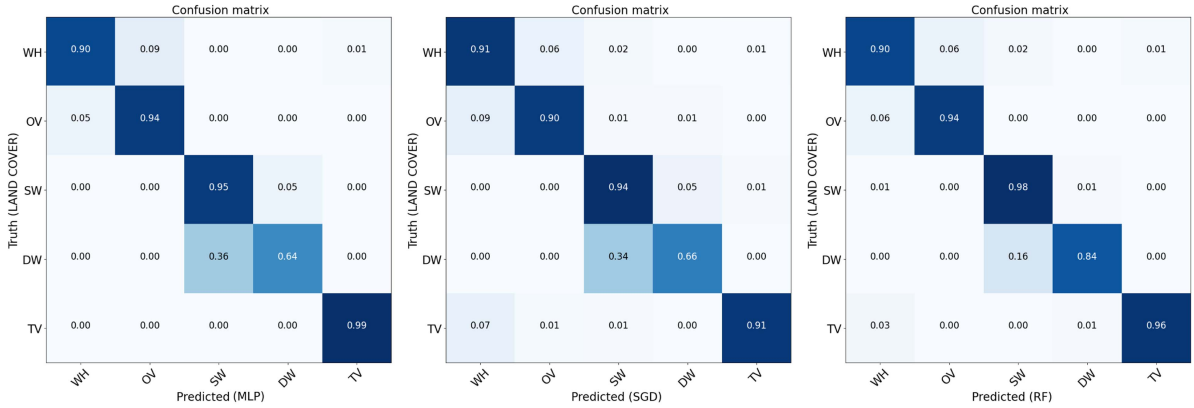

Fig. 5. Confusion matrix for WH mapping and discrimination, where WH = water hyacinth (target class); OV = Other aquatic vegetation; SW= shallow water; DW = deep water; and TV= terrestrial vegetation. (Left) MLP. (Middle) SGD. (Right) RF.

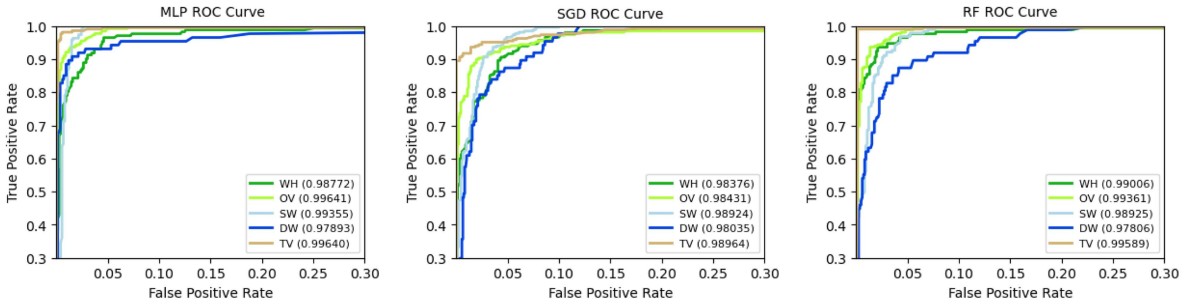

Fig. 6. ROC curves of the three classifiers: (left) MLP, (middle) SGD, and (right) RF, represented as "one versus rest" for each class in the dataset. Numbers in brackets are the AUC values. The ROC curve of the target class, i.e., WH, is shown in dark green.

in Fig. 7, which illustrate how each algorithm classifies the different classes in the image.

The confusion matrix revealed that the RF algorithm could correctly classify the WH class with a high accuracy of 90%. The RF algorithm also performed well in classifying the other classes, with accuracies of 94% for OV, 98% for SW, 84% for DW, and 96% for TV. However, MLP and SGD algorithms showed lower accuracy in detecting WH and distinguishing SW and DW. As explained in Section I, accurate recognition and separation of these areas are essential due to the impact of WH on the water height of the wetland. The ROC curve analysis demonstrated that the RF algorithm had the highest AUC value of 0.99, indicating its superior performance in classification accuracy. In contrast, the MLP and SGD algorithms had lower AUC values of 0.987 and 0.983, respectively. Regarding the IoU accuracy, WH had the highest IoU accuracy of 91%, while the other classifications had the lowest IoU accuracy of 82.3% and 80.1%, respectively.

RF's ability to avoid overfitting and be more robust to missing data and outliers makes it a suitable algorithm for this task. MLP and SGD did not perform as well as RF due to their respective limitations, such as lower resistance to noise and missing data. Although the training data and workflow were the same for all three models, it is important to note that relying solely on results may not provide a complete picture of the model's performance. In this case, despite using the same data and workflow, the SGD algorithm showed errors in the TV in the southeast of AIW compared to Fig. 7.

This highlights the importance of visually comparing the classification maps with the RGB image and in situ field data. By comparing them, we can gain a deeper understanding of their strengths and limitations. This analysis allows us to identify areas where the model may struggle. Overall, the results indicate that the RF algorithm is the most effective ML algorithm for accurately detecting and classifying WH and other vegetation classes from remote sensing data. This matter can provide valuable information for monitoring invasive species in aquatic ecosystems.

B. *Phase II: Modeling WH Growth Stages*

1) Spectral Signature of WH Growth Stages: Our research has yielded fascinating insights into the spectral signature of WH during its various growth stages and its NDVI of sample points of our field visits. Through spectral signature analysis and NDVI calculations, we have confirmed that the data can effectively detect the differences in WH growth stages depicted in the image. This provides a solid foundation for utilizing ML models to accurately identify and classify these variations. Building upon this analysis, we proceeded to separate each stage of WH growth by labeling them. As presented in Fig. 8(a), our investigation into WH's spectral signature revealed that

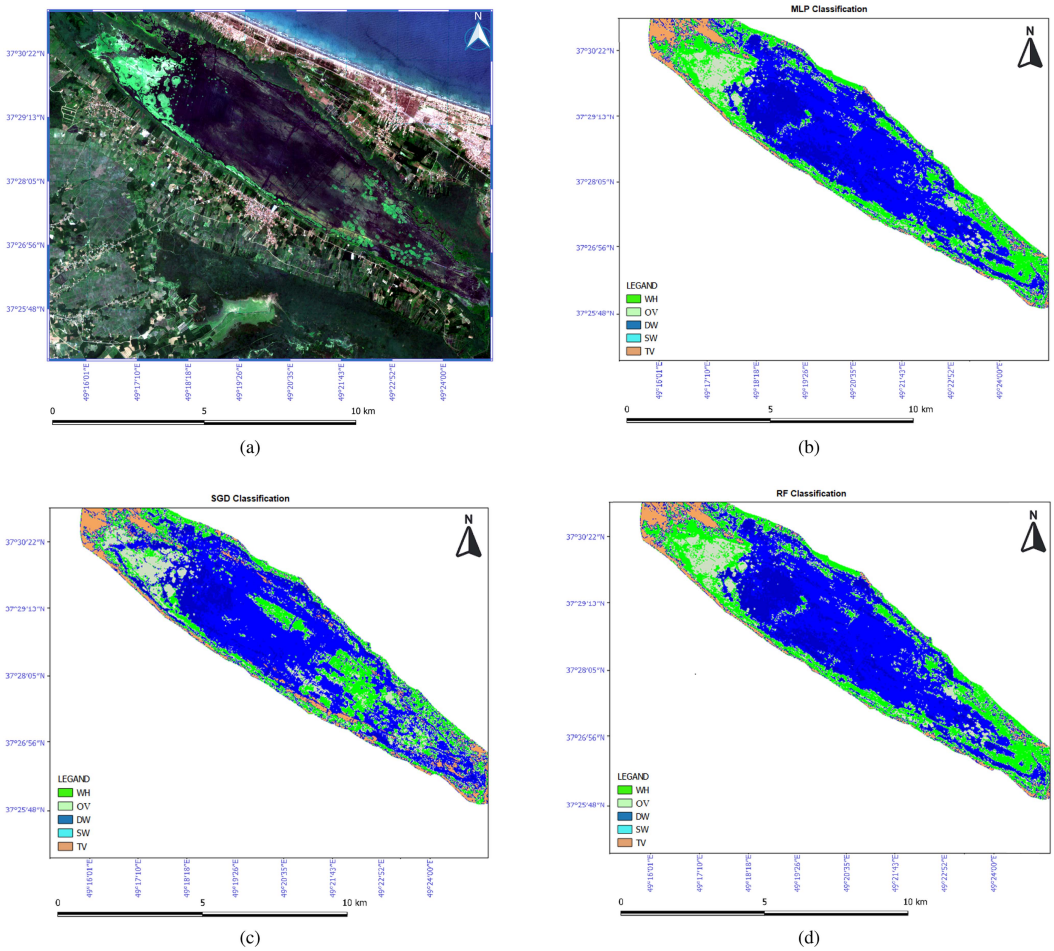

Fig. 7. (a) RGB composite of S2 image acquired on 14 June 2022 over AIW and classification maps of three different ML classifiers: (b) MLP, (c) SGD, and (d) RF.

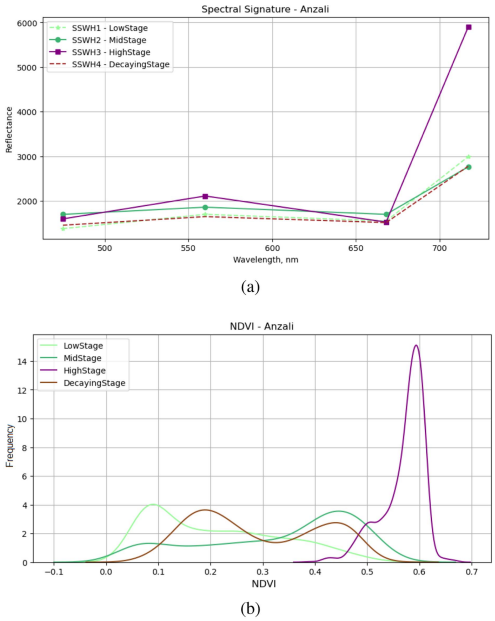

Fig. 8. (a) Spectral signature of WH growth stages for S2 spectral bands (B2, B3, B4, and B8). (b) Kernel density of NDVI plots of WH growth stages.

the plant exhibits higher chlorophyll content during the growth phase, leading to increased reflectance in the NIR regions of the electromagnetic spectrum.

We also computed the kernel density of the NDVI for the sampling points. The probability distributions of the WH growth stages for the NDVI values of sampling points are shown in Fig. 8(b). The higher NDVI values are clearly observed during the high stage. In contrast, during the decaying stage, the chlorophyll content decreases, resulting in reduced reflectance in these regions and lower NDVI values, indicating the presence of a decaying plant. The behavior of WH during the low stage, where a greater difference in reflectance due to higher chlorophyll content is expected, is associated with the plant's youth. One possible explanation for this is saline water. Salty water in the AIW can impose stress on plants, including WH. Salt stress can disrupt nutrient uptake, water balance, and metabolic processes, leading to alterations in chlorophyll synthesis and reflectance properties. The impact of saline water on WH during the low stage can diminish the expected differences in chlorophyll content and reflectance, as the plant may allocate resources toward coping with the salt stress rather than maximizing chlorophyll production. Another possible explanation is the utilization of the

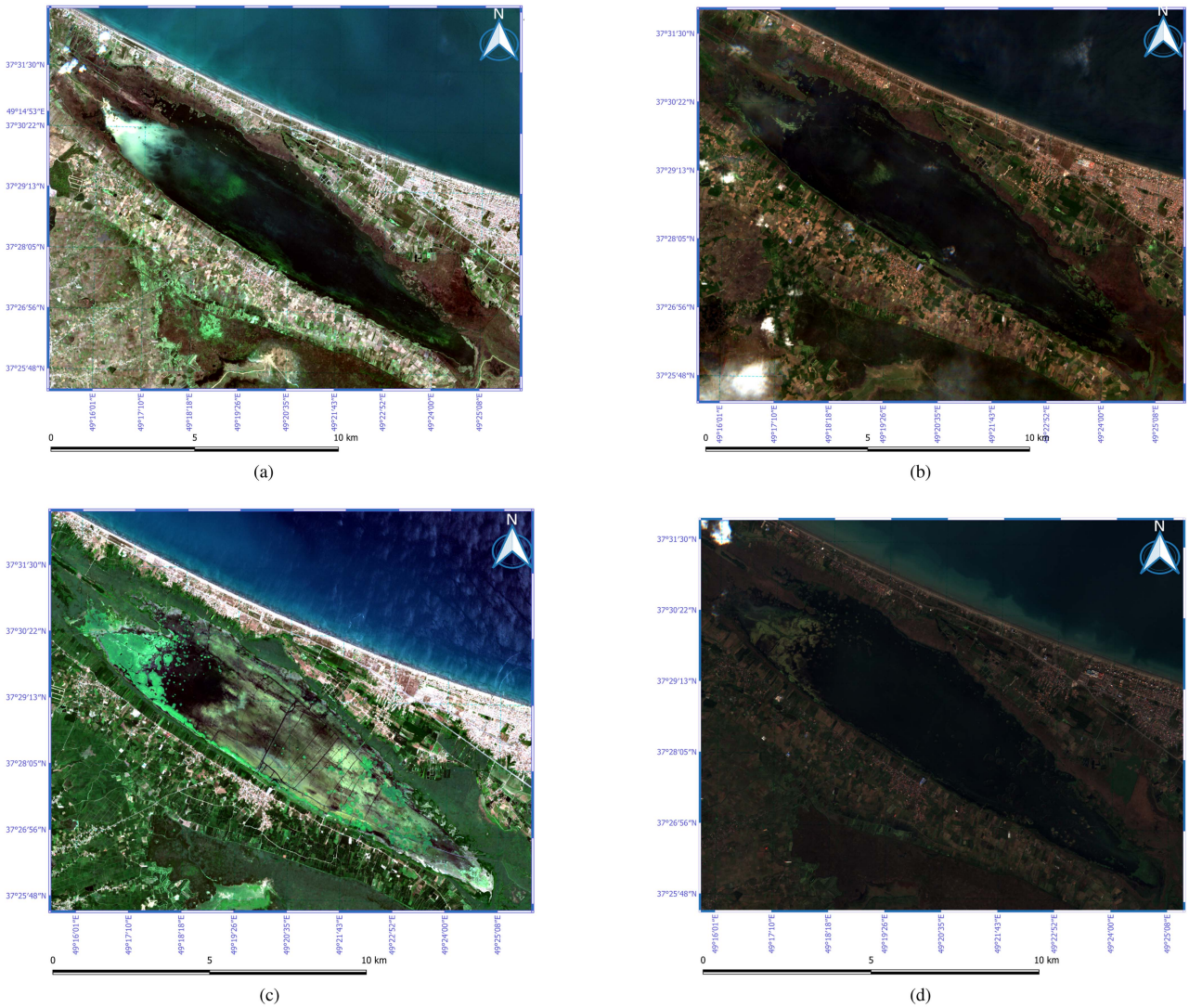

Fig. 9. RGB composite of S2 images used in phase II on four different dates: (a) 31 March 2022, (b) 25 April 2022, (c) 14 July 2022, and (d) 26 November 2022. Images (b) and (d) appear darker due to the sunrise capture time.

S2 Multispectral Instrument satellite data. Using a hyperspectral sensor would be beneficial to analyze the spectral signature in greater detail. Hyperspectral data provide a higher spectral resolution, allowing for a more precise characterization of the spectral properties of WH during different growth stages.

These findings highlight the potential of leveraging spectral data and NDVI measurements to detect and infer WH growth stages, which can have significant implications for managing and controlling invasive species.

Moreover, the insights gained from our research shed light on the physiological changes that occur during the growth cycle of WH. In line with the previous phase results, which identified the RF algorithm as the best choice for classification, in this phase, we employed the RF for the detection and inference of WH growth stages. Applying the RF algorithm enhances our ability to classify and infer WH growth stages accurately.

In summary, our research demonstrates the importance of spectral signature analysis, NDVI calculations, and classification processing in detecting and separating the growth stages of WH depicted in the image. These steps lay the groundwork for further analysis and provide valuable insights into the physiological dynamics of WH. By combining these techniques, we achieve accurate classification and inference, contributing to a deeper understanding of WH growth patterns and supporting effective management strategies for invasive species.

2) Detection and Inference: For the modeling of WH growth stages, our study utilized a dataset consisting of four images, as shown in Fig. 9, each representing at least two WH growth stages at each time point. The dataset comprised a total of 5100 samples, with 4080 samples used for training the RF model and 1020 samples for testing and validation. This dataset division allowed for accurate classification of the different growth stages of WH and reliable evaluation of the model's performance. We trained a RF model with 500 estimators and a depth of 5 using spectral bands and corresponding labeled data. As listed in Table III, the RF algorithm demonstrated an impressive overall accuracy of 98.9%. In addition, the IoU accuracy

TABLE III
CLASSIFICATION RESULTS OF RF CLASSIFIER

Class	F1 (%)	Recall (%)	Precision (%)	IoU (%)
WH1	99.0	98.0	100.0	98.0
WH2	97.8	98.0	97.5	95.6
WH3	98.2	98.5	97.8	96.4
WH4	96.1	95.8	96.3	92.4
Overall accuracy: 98.9%, Kappa coefficient: 0.967				

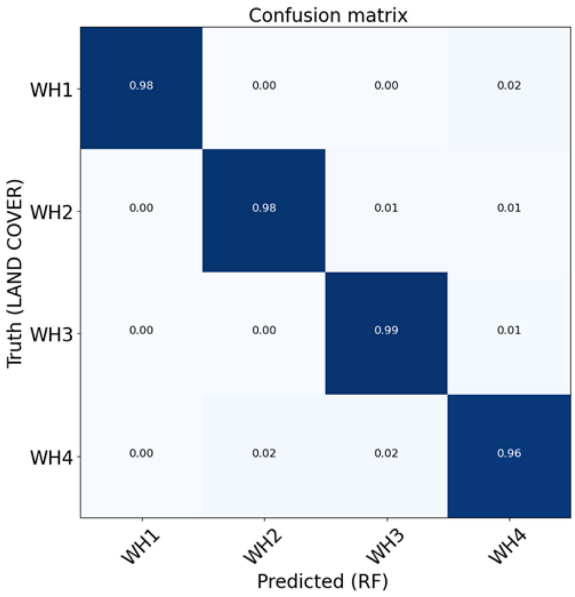

Fig. 10. RF confusion matrix of WH growth stages, where WH1 = WH in early stage, WH2 = WH in mid stage, WH3 = WH in high stage, and WH4 = WH in decaying stage.

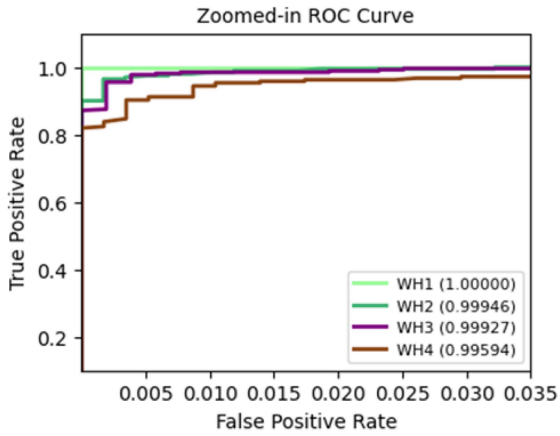

Fig. 11. ROC curves of WH growth stages, represented as "one versus rest" for each class. Numbers in brackets are the AUC values.

was consistently high for all classes, ranging from 92.4% to 98.0%. To evaluate the accuracy of the classification model, we employed a confusion matrix and ROC curve analysis. Fig. 10 shows the confusion matrix demonstrated that the RF algorithm accurately detected and classified each WH growth stage based on their unique spectral signatures in a single image, with an impressive accuracy of over 96% for each class. Fig. 11 displays the ROC curve analysis provided a graphical representation of the tradeoff between true positive and false positive rates for different classification thresholds. The RF algorithm exhibited exceptional performance, with an outstanding AUC value of 0.999, indicating its superior classification accuracy. Altogether, the ML methodology employing RF algorithm achieved a high classification accuracy in inference and classifying the different growth stages of WH from S2. The predictive model can be used to monitor the growth stages of WH to prevent early spread.

The analysis of the output maps in Fig. 12 illustrating the growth stages of WH offers valuable insights into the significance of these visual representations. To gain a deeper understanding of the importance of these output maps, we provide an interpretation. In Fig. 12(a), we can clearly observe the presence of both the low and the decaying stages. The decaying stage signifies the remnants of plants from the previous year. Ideally, effective mitigation efforts would have aimed to minimize the extent of the decaying stage in this image, underscoring the significance of implementing successful control measures against WH to induce an alternative growth cycle. As can be seen in Fig. 12(b), there are small patches of the mid growth stage in the southern region of the wetland. This occurrence may be attributed to the elevated humidity prevalent in the study area, coupled with the influence of urban sewage, which accelerates the growth of WH vegetation. The decaying stage has either transitioned into the low stage or has been displaced by wind to the southeast, outside the water area of AIW. It is plausible that the remaining decaying stage has been consumed by avian species due to the rich biodiversity of the AIW habitat herbivorous fish. The low stage, which was initially observed in the first image, has advanced to the mid stage. Certain portions have been carried southeastward by the prevailing winds, while the rest continues to persist in the northwest. It is of utmost importance to implement prompt measures to prevent the progression of the mid stage to the high stage, as a crucial component of an effective WH management strategy. Consequently, the identification and utilization of this specific stage in Fig. 12(b) become increasingly significant.

In Fig. 12(c), which holds significant scientific value, we can observe all four stages of WH growth. This comprehensive representation underscores the potential for WH to permeate the entire ecosystem when adequate control measures are lacking. As WH reaches the flowering stage, its detrimental impact on the aquatic indigenous wildlife becomes evident, resulting in their demise due to the emission of mustard gas by the plant. It is noteworthy that WH plays a substantial role in the desiccation of AIW. Over the course of our year-long study, we observed a significant decrease in the average wetland depth from 150 to 50 cm, with WH being one of the contributing factors to this drying phenomenon. Fig. 12(d) showcases three stages of WH growth, with the decaying stage being prominently present. The decaying stage signifies the culmination of the WH growth cycle. Without proper intervention, this cycle will persist, leading to the destruction of AIW. Emphasizing the significance of taking necessary measures to prevent the recurrence of this cycle, it is imperative to implement effective strategies for the preservation of AIW.

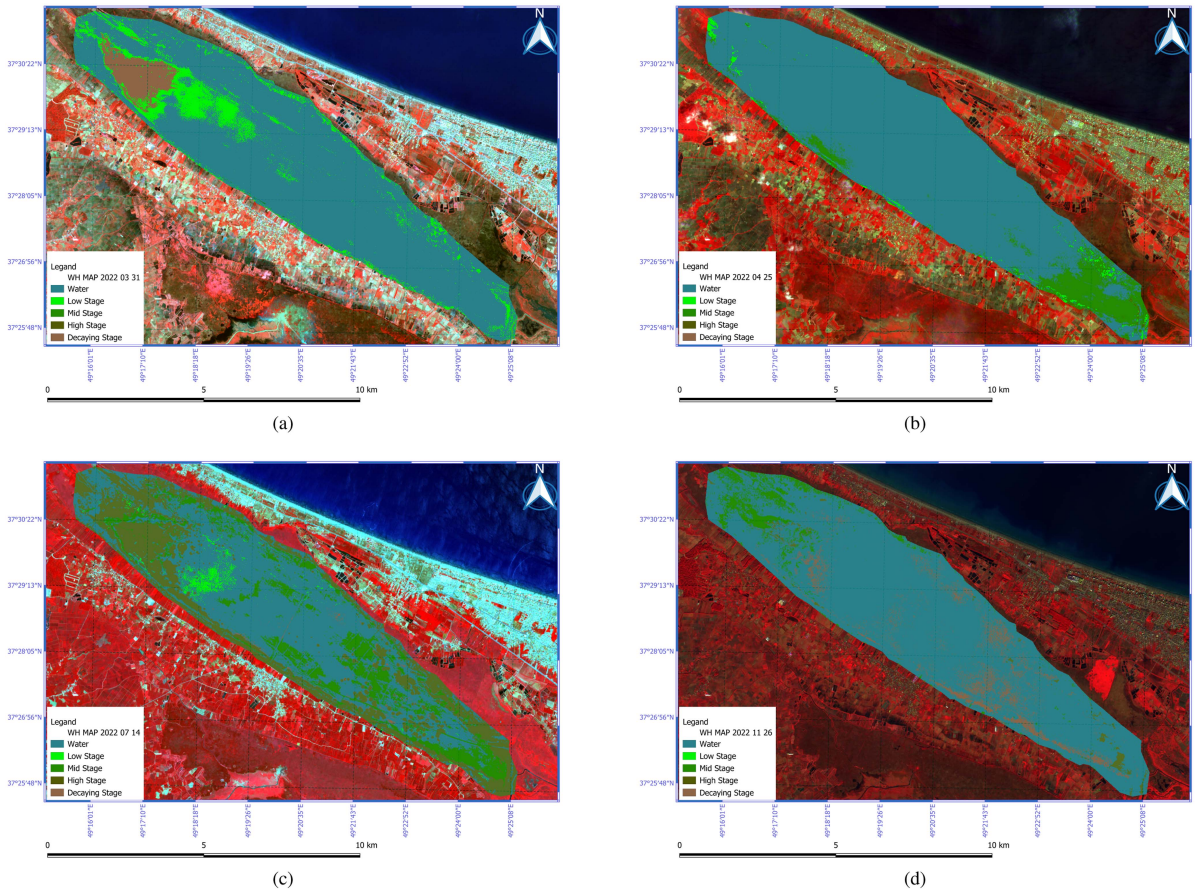

Fig. 12. Detection of WH growth stages on (a) 31 March 2022, (b) 25 April 2022, (c) 14 July 2022, and (d) 26 November 2022 overlaid on false color composite image (NIR, R, G).

IV. Conclusion

Invasive species such as WH can significantly impact aquatic ecosystems, necessitating effective monitoring and control measures. Our study has demonstrated the potential of ML algorithms and remote sensing data for accurately detecting and monitoring the growth stages of WH in aquatic ecosystems.

Although CNN-based methods have demonstrated strong performance in various computer vision tasks, we chose not to incorporate them in this study due to several reasons. First, we focused on exploring alternative approaches that could provide complementary insights into WH growth stages. Second, CNN-based methods are pixelwise, meaning that they operate at the individual pixel level, which often requires a large amount of labeled data for training. However, obtaining such labeled data for WH growth stage analysis, especially when dealing with different plant species coexisting, like AIW, can be challenging. Therefore, we relied on in situ field data to ensure accurate labeling. Finally, by adopting non-CNN techniques, we aimed to explore the feasibility of simpler and more interpretable methods for WH growth stage detection and inference.

We utilized freely available S2 imagery and the RF algorithm to achieve high accuracy in classifying WH infestations and distinguishing them from other classes. Our methodology offers an efficient and cost-effective technique for monitoring the spread of WH, which can have significant ecological and economic impacts on wetland ecosystems.

Our findings also highlight the importance of monitoring the growth stages of WH. The early and mid-growth stages have distinctive spectral signatures, which differ significantly from the decaying stage. These spectral differences have allowed our model to accurately predict the growth stage of WH in a given image, providing crucial insights into the phenology and spread of this invasive species.

One of the most significant contributions of our study is the potential for proactive management and control measures. Our model outputs maps of each growth stage separately, enabling decision makers to take timely and targeted measures to prevent the spread of WH before entering the flowering or high stage, which is the most challenging stage to control.

Our approach provides a valuable environmental monitoring and management tool, enabling more effective control and management of invasive species in aquatic ecosystems. Using ML algorithms and remote sensing data provides a powerful approach for accurately detecting and monitoring invasive species, which can aid in conserving aquatic ecosystems and protecting their biodiversity.

Several challenges emerged in the pursuit of operational implementation. Notably, the growth speed of WH is linked with pollution levels within the lake and other environmental factors. Our data collection efforts were primarily concentrated on a single study site and encompassed only one growth cycle. Furthermore, cloud cover and unfavorable weather conditions frequently obscure satellite imagery, leading to the unusability of certain in situ data when combined with less cloudy S2 data.

A concerted effort is required to expand our dataset to mitigate these challenges and enhance the accuracy of predictive analyses. This expansion necessitates collecting additional in situ data over multiple growth cycles, across diverse lakes and wetlands, and under varying climate conditions. Such an endeavor would fortify the reliability and generalizability of our models, facilitating more effective operational monitoring systems.

A limitation concerns the accessibility of certain WH populations within the AIW. Instances where WH plants are located distant from the wetland boundary present logistical challenges for data collection endeavors. Despite our best efforts, accessing these remote areas proved impractical, constraining our dataset's breadth and representativeness. Moreover, the labor-intensive nature and associated costs of field data collection present additional constraints. WH growth dynamics exhibit rapid fluctuations, necessitating frequent data collection to capture the full spectrum of growth stages accurately. Addressing these limitations requires recognizing the inherent tradeoffs between data comprehensiveness, practicality, and cost-effectiveness. Future endeavors should navigate these challenges effectively, leveraging methodologies to overcome existing constraints and enhance the efficacy of WH monitoring efforts.

In conclusion, our study provides valuable insights into using ML algorithms and remote sensing data for accurate detection and monitoring of the growth stages of WH in AIW. The developed methodology using S2 imagery and RF algorithm can be applied to other areas and ecosystems for monitoring WH and other invasive species. Our findings have exciting implications for future research and for developing more effective monitoring techniques for WH in aquatic ecosystems.

ACKNOWLEDGMENT

The authors would like to sincerely thank Dr. Saeid Homayouni, Institut national de la recherche scientifique, for his invaluable guidance and support throughout this research project.

REFERENCES

[1] R. N. Mack, D. Simberloff, W. M. Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz, "Biotic invasions: Causes, epidemiology, global consequences, and control," *Ecol. Appl.*, vol. 10, no. 3, pp. 689–710, 2000.

[2] D. Pimentel, R. Zuniga, and D. Morrison, "Update on the environmental and economic costs associated with alien-invasive species in the United States," *Ecol. Econ.*, vol. 52, no. 3, pp. 273–288, 2005.

[3] P. Angienda, H. J. Lee, K. Elmer, R. Abila, E. Waindi, and A. Meyer, "Genetic structure and gene flow in an endangered native tilapia fish (Oreochromis esculentus) compared to invasive Nile tilapia (Oreochromis niloticus) in Yala swamp, East Africa," *Conservation Genet.*, vol. 12, no. 1, pp. 243–255, 2011.

[4] N. M. M. Mitan, "Water hyacinth: Potential and threat," *Mater. Today, Proc.*, vol. 19, pp. 1408–1412, 2019.

[5] R. Zarkami, J. Esfandi, and J. Sadeghi, "Modelling occurrence of invasive water hyacinth (Eichhornia crassipes) in wetlands," *Wetlands*, vol. 41, 2021, Art. no. 8.

[6] Ramsar Sites Information Service, "Ramsar convention," 1971. Accessed: 13 May, 2023. [Online]. Available: https://rsis.ramsar.org/ris/40

[7] J. Sadeghi, M. Forio, L. Ho, and P. Goethals, "Evidence-based management of the Anzali wetland system (northern Iran) based on innovative monitoring and modeling methods," *Sustainability*, vol. 13, 2021, Art. no. 5503.

[8] J. Ritchie, P. Zimba, and J. Everitt, "Remote sensing techniques to assess water quality," *Photogrammetric Eng. Remote Sens.*, vol. 69, pp. 695–704, 2003.

[9] M. Guo, J. Li, C. Sheng, J. Xu, and L. Wu, "A review of wetland remote sensing," *Sensors*, vol. 17, 2017, Art. no. 777.

[10] A. Datta et al., "Monitoring the spread of water hyacinth (Pontederia crassipes): Challenges and future developments," *Front. Ecol. Evol.*, vol. 9, 2021, Art. no. 631338.

[11] H. Jafarzadeh, M. Mahdianpari, E. Gill, B. Brisco, and F. Mohammadimanesh, "Remote sensing and machine learning tools to support wetland monitoring: A meta-analysis of three decades of research," *Remote Sens.*, vol. 14, 2022, Art. no. 6104.

[12] I. Iqbal, H. Balzter, F.-E. Bareen, and A. Shabbir, "Identifying the spectral signatures of invasive and native plant species in two protected areas of Pakistan through field spectroscopy," *Remote Sens.*, vol. 13, 2021, Art. no. 4009.

[13] P. Singh et al., *Hyperspectral Remote Sensing in Precision Agriculture: Present Status, Challenges, and Future Trends*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 121–146.

[14] Y. Ghoussein, G. Faour, A. Fadel, J. Haury, H. Abou-Hamdan, and H. Nicolas, "Hyperspectral discrimination of Eichhornia crassipes covers, in the red edge and near infrared in a Mediterranean river," *Biol. Invasions*, vol. 25, pp. 3619–3635, 2023.

[15] S. Steele-Dunne, H. McNairn, A. Monsivais-Huertero, J. Judge, P. Liu, and K. Papathanassiou, "Radar remote sensing of agricultural canopies: A review," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2249–2273, May 2017.

[16] M. Simpson et al., "Detecting water hyacinth infestation in Kuttanad, India, using dual-POL Sentinel-1 SAR imagery," *Remote Sens.*, vol. 14, 2022, Art. no. 2845.

[17] G. Gemechu, X. Rui, and H. Lu, "Wetland change mapping using machine learning algorithms, and their link with climate variation and economic growth: A case study of Guangling county, China," *Sustainability*, vol. 14, 2021, Art. no. 439.

[18] G. Bayable et al., "Detection of water hyacinth (Eichhornia crassipes) in Lake Tana, Ethiopia, using machine learning algorithms," *Water*, vol. 15, no. 5, 2023, Art. no. 880.

[19] V. Akbari et al., "Monitoring aquatic weeds in Indian wetlands using multitemporal remote sensing data with machine learning techniques," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 6847–6850.

[20] I. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Comput. Sci.*, vol. 2, 2021, Art. no. 160.

[21] C. Ongore, C. Aura, Z. Ogari, J. Njiru, and C. Nyamweya, "Spatial-temporal dynamics of water hyacinth, Eichhornia crassipes (Mart.), other macrophytes and their impact on fisheries in Lake Victoria, Kenya," *J. Great Lakes Res.*, vol. 44, pp. 1273–1280, 2018.

[22] A. I. Dogliotti, J. I. Gossn, Q. Vanhellemont, and K. G. Ruddick, "Detecting and quantifying a massive invasion of floating aquatic plants in the Río de la Plata turbid waters using high spatial resolution ocean color imagery," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1140.

[23] C. Nyamekye et al., "Evaluating the spatial and temporal variations of aquatic weeds (biomass) on lower volta river using multi-sensor landsat images and machine learning," *Heliyon*, vol. 7, 2021, Art. no. e07080.

[24] G. Godana, F. F. Feyessa, and G. Debesa, "Eichhornia crassipes expansion detection using geospatial techniques: Lake Dambal, Oromia, Ethiopia," *Environ. Challenges*, vol. 9, 2022, Art. no. 100616.

[25] 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," *Appl. Geography*, vol. 84, pp. 11–22, 2017.

[26] J. d. Mukarugwiro, S. Newete, E. Adam, F. Nsanganwimana, K. Abutaleb, and M. Byrne, "Mapping spatio-temporal variations in water hyacinth (Eichhornia crassipes) coverage on Rwandan water bodies using multispectral imageries," *Int. J. Environ. Sci. Technol.*, vol. 18, pp. 275–286, 2020.

[27] 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. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 228–241, 2023.
[28] E. Rodríguez-Garlito, A. Paz-Gallardo, and A. Plaza, "Mapping invasive aquatic plants in Sentinel-2 images using convolutional neural networks trained with spectral indices," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2889–2899, 2023.
[29] L. Liu et al., "Comparison of machine learning methods applied on multi-source medium-resolution satellite images for Chinese pine (Pinus tabulaeformis) extraction on Google Earth Engine," *Forests*, vol. 13, no. 5, 2022, Art. no. 677.
[30] A. Mirzajani, S. Naderi, and D. Moghadam, "Distribution survey and some biological aspects of water hyacinth in Anzali wetland, Guilan province," *Iranian J. Plant Biol.*, vol. 11, no. 2, pp. 51–62, 2019.
[31] L. Pádua et al., "Water hyacinth (Eichhornia crassipes) detection using coarse and high resolution multispectral data," *Drones*, vol. 6, 2022, Art. no. 47.
[32] K. Thamaga, "Understanding seasonal dynamics of invasive water hyacinth (Eichhornia crassipes) in the greater Letaba river system using Sentinel-2 satellite data," *GISci. Remote Sens.*, vol. 56, pp. 1355–1377, 2019.
[33] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.

Mehran Alizadeh Pirbasti (Student Member, IEEE) received the B.Sc. degree in surveying engineering technology from the Academic Center for Education, Culture and Research, Rasht Institute of Higher Education, Rasht, Iran, in 2020, and the M.Sc. degree in remote sensing engineering from the Hekmat Institute of Higher Education, Qom, Iran, in 2023. He is currently working toward the Ph.D. degree in computer science with University College Dublin, Dublin, Ireland.

His research interests include analyzing different remote sensing datasets from different sensors through artificial intelligence techniques for urban and environmental applications.

Vahid Akbari (Senior Member, IEEE) received the Ph.D. degree in physics with a specialization in radar remote sensing and data analytics from the University of Tromsø(UiT)—The Arctic University of Norway, Tromsø, Norway, in 2013.

In 2014, he performed research in radar remote sensing and machine learning as a Postdoctoral Research Fellow with the UiT—The Arctic University of Norway, the Norwegian Institute of Bioeconomy Research, Bodø, Norway, and the University of Stirling, Stirling, U.K. He was a Visiting Scientist with the Signal Processing and Telecommunications Laboratory, Department of Electrical, Electronic, Telecommunications Engineering and Naval Architecture, University of Genoa, Genoa, Italy, in 2011, and a Visiting Researcher with the German Geoscience Centre, Potsdam, Germany, in 2008. He was an Assistant Professor with the University of Tehran, Tehran, Iran, in 2015. Since 2023, he has been a Lecturer (Assistant Professor) in Data Science/Artificial Intelligence with the University of Stirling. His primary research interests include the intersection of radar remote sensing and statistical modeling/machine learning, with a particular focus on the applications in environmental monitoring.

Deepayan Bhowmik received the Ph.D. degree in electronic and electrical engineering from the University of Sheffield, Sheffield, U.K, in 2011.

He is a Senior Lecturer (Associate Professor) in Data Science and the Director of Business, Innovation and Skill of the School of Computing, Newcastle University, Newcastle Upon Tyne, U.K. His research interests include image processing and computer vision, vision system hardware and optimizations, and multimodal signal and data processing, with a focus on applications in earth observations for sustainability.

Dr. Bhowmik is a recipient of a prestigious Engineering and Physical Sciences Research Council (EPSRC) Dorothy Hodgkin Postgraduate Award. He has been successful in receiving substantial external funds through the U.K. Research Councils (EPSRC and Natural Environment Research Council), Royal Academy of Engineering, EU Horizon 2020, U.K. Defence and Security Accelerator, and industry. His research was showcased at EPSRC Digital Economy Day in 2019, and he is one of the lead editors of the upcoming JPEG Trust standard, which expects to positively impact billions of users across the world.

Savitri Maharaj received the B.Sc. degree in mathematics and computer science from University of the West Indies, Trinidao and Tobago, in 1988 and the M.Sc. degree in computer systems engineering, and the Ph.D. degree in computer science both from the University of Edinburgh, in 1990 and 1996, respectively.

After graduation she was an ERCIM Postdoctoral Research Fellow at INRIA (National Institute for Research in Digital Science and Technology), Sophia-Antipolis, France, and at the Rutherford Appleton Laboratory, Didcot UK. In 1997 she joined the Computing Science and Mathematics department at the University of Stirling, where she is now a Senior Lecturer. Her research interests are varied and multidisciplinary and include the use of agent-based simulation for modelling social and economic systems, and the study of educational computer games.

Since 2018, she has collaborated with a diverse group of colleagues to explore the use of technology, including remote sensing, for monitoring invasive weeds in India.

Armando Marino (Member, IEEE) received the M.Sc. degree in telecommunication engineering from the Università di Napoli "Federico II," Naples, Italy, in 2006, and the Ph.D. degree in polarimetric synthetic aperture radar interferometry from the School of Geosciences, University of Edinburgh, Edinburgh, U.K., in 2011.

In 2006, he joined the High Frequency and Radar Systems Department, German Aerospace Centre, Oberpfaffenhofen, Germany, where he developed the M.Sc. thesis. From March 2011 to October 2011, he was with the Institute of Computing Research, University of Alicante, San Vicente del Raspeig, Spain. From 2011 to 2015, he was a Postdoctoral Researcher and Lecturer with the Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland. In June 2015, he was a Lecturer with the School of Engineering and Innovation, Open University, Milton Keynes, U.K. Since May 2018, he has been an Associate Professor with the Faculty of Natural Sciences, University of Stirling, Stirling, U.K.