Geosciences

AIMS Geosciences, 10(1): 62–79.
DOI: 10.3934/geosci.2024005
Received: 15 September 2023
Revised: 24 October 2023
Accepted: 12 November 2023
Published: 25 January 2024

http://www.aimspress.com/journal/geosciences

Research article

Machine Learning-Based Classification of Small-Sized Wetlands Using Sentinel-2 Images

Eric Ariel L. Salas*, Sakthi Subburayalu Kumaran, Robert Bennett, Leeoria P. Willis and Kayla Mitchell

Agricultural Research Development Program (ARDP), Central State University, Wilberforce, OH 45384, USA

* **Correspondence:** Email: esalas@centralstate.edu; Tel.: +01-937-376-6637.

Abstract: Wetlands are invaluable ecosystems, offering essential services such as carbon sequestration, water purification, flood control and habitat for countless aquatic species. However, these critical environments are under increasing threat from factors like industrialization and agricultural expansion. In this research, we focused on small-sized wetlands, typically less than 10 acres in size, due to their unique ecological roles and groundwater recharge contributions. To effectively protect and manage these wetlands, precise mapping and monitoring are essential. To achieve this, we exploited the capabilities of Sentinel-2 imagery and employ a range of machine learning algorithms, including Random Forest (RF), Classification and Regression Tree (CART), Gradient Tree Boost (GTB), Naive Bayes (NB), k-nearest neighbors (KNN) and Support Vector Machine (SVM). Our evaluation used variables, such as spectral bands, indices and image texture. We also utilized Google Earth Engine (GEE) for streamlined data processing and visualization. We found that Random Forest (RF) and Gradient Tree Boost (GTB) outperformed other classifiers according to the performance evaluation. The Normalized Difference Water Index (NDWI) came out to be one of the important predictors in mapping wetlands. By exploring the synergistic potential of these algorithms, we aim to address existing gaps and develop an optimized approach for accurate small-sized wetland mapping. Our findings will be useful in understanding the value of small wetlands and their conservation in the face of environmental challenges. They will also lay the framework for future wetland research and practical uses.

Keywords: GEE; mapping wetlands; remote sensing wetlands; random forest; Google Earth Engine

1. Introduction

Among important ecosystems, wetlands are unique. These ecosystems are useful for hydrological cycle, carbon sequestration, water purification, flood control and provide habitats for thousands of aquatic flora and fauna [1]. Wetlands are, however, threatened by industrialization and agricultural intensification [2]. In the USA, the main cause of wetland loss is the conversion to agricultural land use [3]. Globally, about 35% of wetlands have been lost between 1970 and 2015 [4].

Small-sized wetlands, the focus of our study, comprise a diverse array of ecologically and hydrologically significant ecosystems. Typically occupying an area of less than 10 acres (0.04 km^2), these wetlands are situated within a variety of terrestrial and aquatic landscapes [5]. Hydrologically, these wetlands are noticeable by their remarkable ability to retain water over extended periods, particularly during wet seasons, thereby contributing substantially to local groundwater recharge [6]. From an ecological perspective, small-sized wetlands hold a fundamental role in supporting unique habitats for a diverse range of wetland-dependent species [7]. These types of wetlands usually have specific vegetation types, such as emergent aquatic plants and hydric soils, further setting them apart from the surrounding ecosystems [8].

It is crucial to map and monitor wetlands for conservation and management strategies [9]. Remote sensing technology, including satellite imagery, has emerged as a valuable tool for mapping and monitoring wetlands at a regional and global scale [10,11]. The European Space Agency's (ESA) Sentinel satellites, specifically Sentinel-2, have proven particularly useful in providing high-resolution optical and radar imagery to map wetlands. The multispectral Sentinel-2 imagery can identify vegetation and water bodies, while its high-resolution capabilities enable the detection and characterization of small-sized wetlands [12].

To enhance the precision and efficiency of wetland classification using Sentinel-2 imagery, there has been a growing interest in the application of machine learning (ML) algorithms. Machine learning is a subset of artificial intelligence and involves training algorithms to recognize patterns in data and subsequently classify new data based on these learned patterns. In the context of satellite imagery, ML algorithms offer the potential for automating the mapping process and accurately classifying various land cover types, including wetlands [13]. Among the ML algorithms, Random Forest (RF) and Classification and Regression Tree (CART) have emerged as prominent choices for mapping wetlands with Sentinel-2 images [14,15]. RF, for instance, leverages the combination of multiple decision trees to produce precise predictions through the aggregation of their outputs. Its capacity to handle complex and high-dimensional datasets, while also enabling the assessment of variable importance, positions RF as a robust tool for wetland mapping [16]. Similarly, CART constructs binary decision trees based on feature attributes, facilitating the accurate classification of wetland areas [17].

In addition to RF and CART, various other ML algorithms have found application in the mapping of small-sized wetlands using Sentinel-2 imagery. Among these, Gradient Tree Boost (GTB) stands out as an ensemble learning method, similar to RF, which combines multiple decision trees to enhance classification accuracy [18]. Naive Bayes (NB) is another notable algorithm, offering a probabilistic approach that assumes independence between features, and it has demonstrated successful outcomes in wetland mapping [19]. In a different manner, k-nearest neighbors (KNN), a non-parametric algorithm, distinguishes itself by classifying data points based on their proximity to known data points [20]. Further, the Support Vector Machine (SVM) also emerges as an effective ML tool, constructing hyperplanes to delineate distinct classes within the feature space [21].

Several studies have been conducted regarding the effectiveness of these ML algorithms in wetland mapping tasks. For example, Mahdianpari et al. [22] and Waleed et al. [23] employed RF and CART algorithms with Sentinel-2 images to achieve high overall accuracies (>95%) in wetland classification. Li et al. [24] utilized GTB for wetland classification and reported improved accuracy compared to individual decision trees. Wu et al. [25] successfully applied NB and KNN algorithms in wetland mapping and highlighted its ability to handle complex classification tasks. Finally, Gemechu et al. [26] demonstrated the effectiveness of SVM algorithm in temporal wetland classification in Guangling County, China with accuracies ranging from 86% to 98.1%.

To process and analyze the Sentinel-2 images using these ML classifiers, we take advantage of the capabilities of Google Earth Engine (GEE), a cloud-based platform that combines a vast archive of satellite imagery with geospatial analysis functions [27]. GEE provides a powerful computing infrastructure and pre-built functions for image analysis and enables efficient data processing and visualization. It offers access to Sentinel satellite imagery, including Sentinel-2 data and facilitates the integration of ML algorithms for wetland mapping [28,29]. GEE could overcome computational limitations and process various satellite imagery at the same time, which allows for the integration of multiple ML algorithms and the exploration of various input features for wetland classification [27].

Despite the advancements in ML algorithms and the utilization of Sentinel-2 imagery for wetland mapping, there is a research gap in the development of an integrated approach that combines multiple machine learning algorithms to improve the accuracy and efficiency of small-sized wetland classification. While studies have individually demonstrated the effectiveness of algorithms, especially RF, there is a need to explore the potential of other algorithms in a synergistic manner to enhance wetland mapping processes. Our specific objectives of this study include:

1. Assessing the individual performance of RF, CART, GTB, NB, kNN and SVM classifiers in wetland classification using Sentinel-2 imagery.
2. Evaluating the impact of different input variables, such as spectral bands from Sentinel-2 images, spectral indices and image texture on the performance of each classifier.
3. Mapping of the wetland areas using the best performing classifier.

Our results would provide valuable insights into the development of an optimized and accurate wetland mapping approach. By exploring the synergistic potential of these algorithms and variables, our research aims to address existing gaps and develop an optimized approach for accurate small-sized wetland mapping. Such an approach would be helpful for other researchers, conservationists, land managers, policy-makers and environmental practitioners who are actively involved in the preservation and management of small-sized wetlands. Finally, our results would also address the gaps that exist in the National Wetlands Inventory (NWI) database.

2. Materials and methods

2.1. Study Area

The study area includes a number of wetland systems, including floodplain woodlands, marshes, wet prairies, woodland pools and scrub-shrub wetlands. These small-sized wetlands are managed by Beaver Creek Wetlands Greenway (BWG) Community Land Trust. Located in Beavercreek, Ohio, USA, the wetlands are protected by the Ohio Environmental Protection Agency (OEPA) Clean Water Act. Due to the variety of flora and fauna species, school groups, birdwatchers and wildflower enthusiasts frequent the wetlands. The Beaver Creek stream runs through the wetlands from north to south. The BWG boundary and the perimeter of our study area are depicted in Figure 1. According to the NLCD land use, most of the urban areas are in the east, while cultivated crops predominate in the west. Although wetland is one of the NLCD classes, due to their small sizes, they are not properly mapped out.

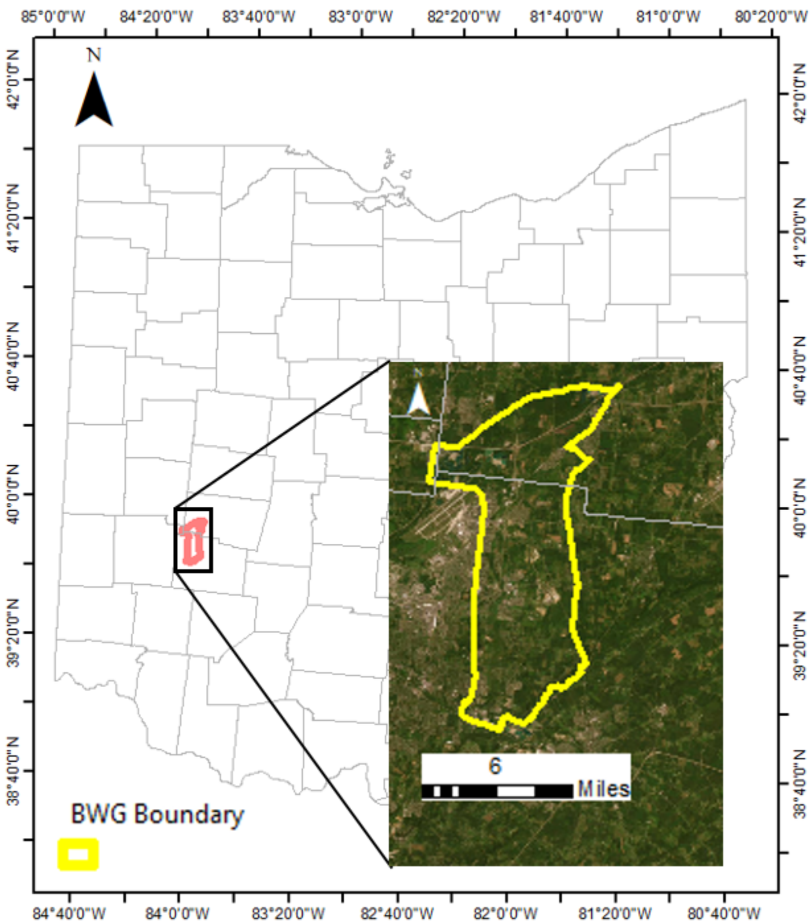

Figure 1. The BWG study area is located in southwestern Ohio.

2.2. General Methodology

The methods for classifying wetlands using GEE and utilizing six classifiers involved several steps.

1. Sentinel-2 images were retrieved and preprocessed using GEE. Relevant spectral bands and features were extracted from the images, capturing important information about wetland vegetation and water characteristics.
2. The six machine learning classifiers, including RF, CART, GTB, NB, KNN and SVM, were implemented to classify the wetland areas.
3. These classifiers were trained on the extracted spectral indices, image texture and spectral bands, leveraging the ability of the classifiers to recognize patterns and classify new data based on learned patterns.

4. The performance of each classifier was assessed individually, evaluating their accuracy in wetland classification using Sentinel-2 imagery.
5. The impact of different input variables on the classification performance of each approach was also evaluated by grouping them into four sets: Set A (spectral bands, spectral indices, image texture); set B (spectral bands, spectral indices); set C (spectral bands, image texture); set D (spectral bands).
6. The accuracy of the multiple classifiers using the various set of variables was validated by splitting the sampling data into training and testing, providing insights into an optimized and accurate wetland mapping approach.

2.3. Google Earth Engine (GEE) Platform and Sentinel-2 Images

GEE *(earthengine.google.com)* is a cloud-based platform that provides researchers with a vast collection of satellite imagery and geospatial datasets [30]. It offers a comprehensive and diverse range of data sources, including the 10-meter spatial resolution Sentinel-2 which is essential for wetland mapping. The extensive GEE data archive, coupled with its efficient data storage and processing capabilities, allowed this research seamless access to the necessary image data for analysis. By utilizing the power of GEE, we retrieved and preprocessed the Sentinel-2 imagery from June 1 to October 31, 2022 (summer to fall seasons), extracted relevant spectral bands and features from the images and performed various data manipulations required for wetland classification. The use of GEE ensures a streamlined and efficient workflow, enabling us to leverage its vast data resources and advanced analysis capabilities for accurate wetland mapping.

We utilized summer to fall images for wetland classification because this period captures crucial temporal variations in wetland vegetation dynamics and water presence. During this period, wetlands undergo distinct changes in vegetation phenology, water levels and land cover, which can provide valuable information for accurate classification. By analyzing imagery from June 1 to October 31, 2022, we captured the seasonal transitions and phenological patterns necessary for distinguishing wetland classes and improved the overall accuracy of wetland mapping.

2.4. Selected Classifiers RF, CART, GTB, NB, KNN and SVM

We selected RF, CART, GTB, NB, KNN and SVM as our classifiers based on their demonstrated effectiveness in wetland classification using remote sensing data. Each classifier offers unique advantages and has been widely used in various land cover mapping studies.

RF is an ensemble learning method that combines multiple decision trees to generate accurate predictions by aggregating their outputs [31]. It utilizes the concept of bagging and random feature selection to reduce overfitting and improve generalization. The RF algorithm combines the predictions of individual decision trees using a voting mechanism or averaging method, as shown in Equation 1:

$$\hat{y}RF = \frac{1}{N}\sum_{i=1}^{N} \quad \hat{y}_i(x), where\ \hat{y}_i(x)\ \epsilon\ \{-1,1\} \tag{1}$$

where $\hat{y}RF$ is the predicted class label for a given sample $x, \hat{y}_i, (x)$ is the predicted class label of the i-th decision tree and *N* is the total number of decision trees in the forest.

CART constructs binary decision trees based on feature attributes, enabling precise classification of wetland areas [32]. The splitting criterion used in CART is often based on Gini impurity or information gain, as shown in Equation 2:

$$Gini(p) = 1 - \sum_{k=1}^{K} \quad pk^2 \tag{2}$$

where pk represents the proportion of samples in class k at a given node. CART recursively partitions the feature space based on the selected splitting criterion, forming a binary tree structure.

GTB, like RF, is an ensemble learning method that combines multiple decision trees to improve classification accuracy [32]. GTB differs from RF in the way it constructs subsequent trees. The boosting process of GTB updates the weights of misclassified samples, allowing subsequent trees to focus on the previously misclassified samples. The final prediction of GTB is a weighted sum of predictions from all the trees, as shown in Equation 3:

$$\hat{y}GTB = \frac{1}{N}\sum_{i=1}^{N} \quad v\hat{y}_i(x), where\ \hat{y}_i(x)\ \epsilon\ \{-1,1\} \tag{3}$$

where $\hat{y}GTB$ is the predicted class label for a given sample $x, \hat{y}_i, (x)$ is the predicted class label of the *i*-th decision tree, *N* is the total number of decision trees in the ensemble and *v* is the learning rate.

NB is a probabilistic classifier that assumes independence between features and has been applied successfully in wetland mapping [33]. It estimates the conditional probability of a sample belonging to a particular class using Bayes' theorem. The predicted class label is determined by selecting the class with the highest probability, as shown in Equation 4:

$$\hat{y}NB = argmax_{y\epsilon\{1,2,\ldots,K\}}P(y)\prod_{i=1}^{n} \quad P(x_i|y) \tag{4}$$

where $\hat{y}NB$ is the predicted class label for a given sample, $P(y)$ is the prior probability of class y, $P(y)$ is the conditional probability of feature x_i given class *y*, *n* is the number of features and *K* is the total number of classes.

KNN is a non-parametric algorithm that classifies data points based on their proximity to other known data points [34]. It assigns the class label based on the majority vote of its nearest neighbors. The class label of a sample is determined by Equation 5:

$$\hat{y}KNN = mode(\{y_i|x_i \text{ is one of the } k \text{ nearest neighbors of } x\}) \tag{5}$$

where $\hat{y}KNN$ is the predicted class label for a given sample $x_i y_i$ is the class label of the *i*-th nearest neighbor and *k* is the number of nearest neighbors.

SVM is a powerful algorithm that constructs a hyperplane to separate different classes in the feature space [35]. It aims to maximize the margin between the support vectors and the decision boundary. The class label of a sample is determined by the sign of the discriminant function, as shown in Equation 6:

$$\hat{y}SVM = \ sign\ (f(x)) \tag{6}$$

where $\hat{y}SVM$ is the predicted class label for a given sample $x, f(x)$ is the discriminant function and the sign function determines the class label based on the sign of the discriminant function.

Using the strengths of these classifiers, we aim to improve the accuracy and efficiency of small wetland classification.

2.5. Image Texture and Other Indices as Predictors

We added image texture measures such as entropy, along with vegetation indices like NDVI and water indices like NDWI as they have crucial roles in wetland classification, as they can capture important spectral and textural characteristics of wetland environments. Entropy, as a measure of randomness or disorder within an image, provides valuable information about the spatial distribution and heterogeneity of land cover classes within a wetland area. Texture measures have been widely used in remote sensing studies for land cover classification and change detection [10,36].

NDVI is a widely applied vegetation index that quantifies the presence and vigor of vegetation by leveraging the contrast between near-infrared (NIR) and red reflectance. Wetlands typically exhibit distinct vegetation characteristics, and NDVI helps in differentiating wetland vegetation from other land cover types [14,37]. High NDVI values indicate dense vegetation cover, which is often associated with wetland areas due to their unique hydrological conditions and plant species composition. NDWI is a spectral index commonly used to identify water bodies. It exploits the contrast between the NIR and shortwave infrared (SWIR) bands to highlight the presence of water [21]. Wetlands are characterized by the presence of water, and the inclusion of NDWI as a variable helps in accurately delineating wetland boundaries and differentiating wetlands from non-wetland land cover classes [19].

By incorporating entropy with NDVI and NDWI, together with the Sentinel-2 bands (B2: blue; B3: green; B4: red; B5, B6, B7, B8: VNIR; B11, B12: SWIR) into the classification process, we effectively captured the heterogeneity of wetland environments, distinguish wetland vegetation from other land cover types and accurately identify water bodies within wetland areas. This comprehensive set of variables enhanced the discriminatory power of the classification algorithms and enhanced the accuracy of wetland mapping and monitoring studies.

2.6. Sets of Variables to Assess the Impact of Each Predictor

We ran the classifiers using four different sets of variables (A, B, C and D) to understand the impact of including different types of variables on the classification performance.

1. Set A (spectral bands, spectral indices, image texture): This set includes a comprehensive range of variables, namely spectral bands, spectral indices (vegetation and water) and image texture. Spectral bands capture information about the reflectance properties of different land cover classes, while spectral indices, such as the NDVI and NDWI, provide insights into specific land cover characteristics like vegetation and water content. Image texture measures, such as entropy, describe the spatial arrangement and patterns within the imagery. By combining these variables, the classifiers can leverage a diverse set of information for improved discrimination and classification accuracy.
2. Set B (spectral bands, spectral indices): This set focuses on combining spectral bands and spectral indices, excluding the image texture measures. It allows us to evaluate the contribution of spectral indices in enhancing the performance of the classifiers. Spectral indices, being derived from specific band combinations, provide valuable information about vegetation health, moisture content and other land cover characteristics. By including these indices alongside spectral bands,

the classifiers can utilize additional spectral information for more accurate classification.

3. Set C (spectral bands, image texture): This set explores the impact of image texture measures when combined with spectral bands. Image texture measures capture information related to spatial patterns, such as the heterogeneity or smoothness of land cover classes. By incorporating image texture alongside spectral bands, the classifiers can consider textural characteristics as discriminative features, which may aid in distinguishing between different land cover types.
4. Set D (spectral bands): This set serves as a baseline comparison where only spectral bands are used as input variables. By excluding spectral indices and image texture measures, the classifiers' performance can be assessed based solely on the spectral information captured by the satellite imagery. This set helps evaluate the extent to which additional variables (spectral indices and image texture) contribute to improving the classification accuracy compared to using spectral bands alone.

2.7. Field Data and Accuracy Assessment

We collected 60 points samples from the field (using GPS) and interpreted additional 20 samples from the high-resolution images using Google Earth. To assess the wetland mapping classification accuracy of the classifiers, we employed a data splitting approach to create training and testing datasets. We divided our dataset into 70% training set and 30% testing set. This division of dataset guarantees an unbiased assessment of the classifiers' accuracy by evaluating their ability to generalize to new data.

During the training phase, the classifiers were trained using the labeled training dataset, where each pixel was assigned a known wetland class label. This process involved learning the underlying patterns and relationships between the spectral characteristics and wetland classes present in the training data. The trained classifiers were then applied to the testing dataset, where the class labels were withheld, and the classifiers' predictions were compared against the true class labels to evaluate their accuracy.

We assessed the accuracy using established metrics such as overall accuracy (OA) and kappa coefficient. These metrics provide quantitative measures of the classifier performance in correctly identifying wetland classes and assessing the agreement between the predicted and reference class labels [38]. By utilizing the training and testing data splitting approach, we evaluated the accuracy of the RF, CART, GTB, NB, KNN and SVM classifiers and compared their mapping performance.

3. Results

3.1. Performance of the Six Classifiers

We ran the classifiers using four different sets. GTB demonstrated high overall training and validation accuracies (from 0.94 to 0.97), indicating good classification performance (Table 1). These numbers suggested a substantial level of agreement between the predicted and reference classifications. The RF classifier exhibited equally high overall training and validation accuracies, ranging from 0.93 to 0.98, and the kappa coefficients ranging from 0.97 to 0.98.

The SVM showed relatively lower overall training and validation accuracies, ranging from 0.90 to 0.92, while kappa coefficients, ranging from 0.86 to 0.91. CART demonstrated high overall training accuracy (0.91 to 0.95), with kappa coefficients ranging from 0.92 to 0.95. KNN showed lower overall training and validation accuracies, both consistently at 0.67. The kappa coefficient was also low at

0.48. Finally, NB exhibited moderate overall training accuracy, ranging from 0.72 to 0.75. The performance in terms of kappa coefficient ranged from 0.72 to 0.73.

Based on the provided results, the RF and GTB classifiers outperformed the other classifiers in terms of OA and kappa coefficients. These two classifiers demonstrated higher accuracy and agreement with the reference classifications, indicating their effectiveness in mapping the LCLU/wetland classes. In the end, we used the RF and GTB classifiers as the final models for mapping the wetlands.

Table 1. Overall training accuracy/validation accuracy/kappa of the six ML classifiers in mapping the LCLU classes. A= spectral bands, spectral indices, image texture; B = spectral bands, spectral indices; C = spectral bands, image texture; and D = spectral bands.

Classifier	A	B	C	D
CART	0.95/0.91/0.95	0.91/0.90/0.92	0.90/0.90/0.92	0.91/0.90/0.92
RF	0.98/0.94/0.98	0.98/0.93/0.98	0.98/0.93/0.97	0.97/0.92/0.97
SVM	0.92/0.91/0.86	0.92/0.90/0.86	0.91/0.90/0.86	0.91/0.90/0.86
GTB	0.97/0.94	0.97/0.93	0.97/0.93	0.97/0.93
KNN	0.67/0.67/0.48	0.67/0.66/0.48	0.67/0.65/0.48	0.67/0.67/0.48
NB	0.75/0.73	0.74/0.73	0.75/0.72	0.72/0.73

When we evaluated the efficacy of the classifiers in mapping wetlands against all other classes (that is, excluding the wetland class), the results indicated varying levels of relative OA (Table 2). Among the classifiers tested, RF and GTB exhibited the highest relative OA values of 0.93, suggesting their effectiveness in accurately mapping wetlands against the combined class. Both RF and GTB outperformed other classifiers in correctly classifying the samples.

Table 2. Overall training accuracy/validation accuracy/kappa of the six ML classifiers in mapping wetlands against all other classes. A= spectral bands, spectral indices, image texture; B = spectral bands, spectral indices; C = spectral bands, image texture; and D = spectral bands.

Classifier	Relative Overall Accuracy (Wetlands)
CART	0.90
RF	0.93
SVM	0.90
GTB	0.93
KNN	0.66
NB	0.73

CART and SVM achieved a relative OA of 0.90, indicating similar performance in accurately classifying the wetlands against the combined class. Although CART and SVM performed slightly lower than RF and GTB, they demonstrated good accuracy. In contrast, KNN exhibited a lower relative OA of 0.66, indicating a comparatively weaker performance in correctly classifying the wetlands. This suggests that KNN may struggle to distinguish wetlands from other land cover types effectively. NB achieved a moderate relative OA of 0.73, indicating its moderate performance. While NB performed better than KNN, it fell behind RF, GTB, CART and SVM in accuracy.

Overall, the results highlighted the superior performance of RF and GTB classifiers in accurately mapping the wetlands, suggesting their suitability for distinguishing and mapping wetlands. CART and SVM classifiers also exhibited good accuracy, while KNN showed relatively weaker performance, and NB achieved moderate accuracy.

3.2. Ranking the Importance of the Variables

We evaluated the importance of specific variables in the classification when using only two classes (wetlands vs others) (Table 3). Across the different variable sets (A, B, C and D), certain variables consistently emerged as significant contributors to the classification process. In variable set A, which includes spectral bands, spectral indices and image texture, B4 (Red spectral band), B11 (SWIR spectral band) and NDVI (spectral index) were identified as the most important variables. Similarly, in variable set B (spectral bands and spectral indices), B11, B8, B5 and B2 (all spectral bands) played crucial roles in distinguishing between the two classes. Variable set C (spectral bands and image texture) further emphasized the significance of B11, B8, B5, B2 and B12 (spectral bands) in accurate classification. Finally, variable set D (spectral bands) confirmed the importance of B11, B8, B3, B2, B4 and B5 (all spectral bands) for effective discrimination between the two classes. Among all variables, entropy is absent from the list of important variables, suggesting that it may have had limited discriminatory power in the classification. Overall, the consistent ranking of these variables underscores their relevance in accurately classifying the land cover classes under consideration, emphasizing the value of specific spectral bands and spectral indices, such as NDVI, in the classification process.

Table 3. Important variables in the classification. A= spectral bands, spectral indices, image texture; B = spectral bands, spectral indices; C = spectral bands, image texture; and D = spectral bands.

Rank	A	B	C	D
1	B4	B11	B11	B11
2	B11	B8	B8	B8
3	NDVI	B5	B5	B3
4	B2	B2	B2	B2
5	B3	B12	B12	B4
6	B12	NDWI	B4	B5

3.3. Mapping the Wetlands Using the Best Performing Classifiers

We mapped the wetlands the two best-performing classifiers (GTB and RF), using Sentinel-2 imagery through the GEE environment (Figure 2).

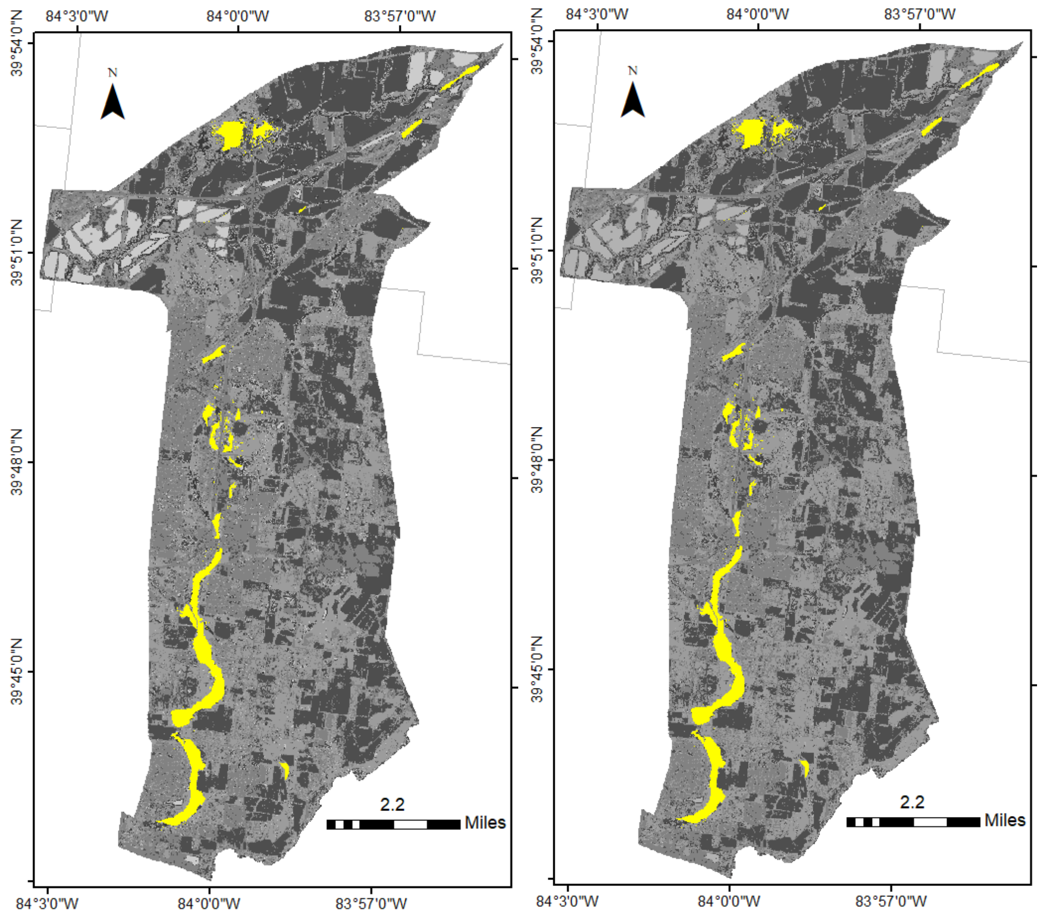

Figure 2. Resulting classified map of the BWG wetlands using the (a) GTB and (b) RF models. The wetland class is highlighted in yellow.

4. Discussion

4.1. Performance of the Six Classifiers

We conducted an assessment of the effectiveness of six popular ML classifiers in the framework of small-sized wetland mapping. Our findings revealed a range of performance levels among the classifiers. Notably, RF exhibited a higher degree of accuracy, followed by CART, SVM and GTB. Conversely, KNN and NB exhibited comparatively lower levels of accuracy. These results agreed with prior research findings that support the efficacy of decision tree-based classifiers for wetland mapping [17,18,20].

In contrast to our results, Gemechu et al. [26] presented findings indicating that SVM classifier outperformed RF and other classifiers in terms of accuracy, achieving levels as high as 98.1% in wetland mapping. They particularly underscored the proficiency of SVM in handling high-dimensional data and capturing complex relationships, which ultimately contributed to improved classification accuracy. Furthermore, a study by Amani et al. [39] reported satisfactory results with the k-nearest neighbors (KNN) classifier when applied to wetland mapping using remote sensing data. Their results placed importance on the significance of optimizing KNN algorithm parameters to achieve precise wetland classification. These divergent results underscore the importance of taking into account specific contextual variables and parameters when making decisions regarding classifier selection for wetland mapping applications. Considerations relating to the study area, dataset characteristics and optimization techniques should be weighed cautiously to ensure dependable and accurate wetland classification results.

Furthermore, Xing et al. [40] conducted a comprehensive investigation of ensemble classifiers, among them RF, in the context of wetland mapping. Their investigation yielded high overall accuracy when using RF; however, they observed a tendency for overestimating wetland areas attributed to inherent algorithmic characteristics. In response, they recommended incorporating post-processing techniques to mitigate this challenge and enhance the precision of wetland mapping.

Our study contributes to the existing pool of knowledge pertaining to wetland mapping by conducting a thorough evaluation of various ML classifiers aimed at mapping small-sized wetlands. Although our findings indicated a relatively high level of accuracy in the case of decision tree-based classifiers, it remains imperative for future research studies to continue exploring the performance of the classifiers within different contextual frameworks. Additionally, the prospect of integrating complementary approaches to advance wetland mapping accuracy warrants further investigation.

4.2. Spatial Distribution of Wetlands

One of our primary objectives was the precise mapping of small-sized wetlands, a task for which conventional approaches, illustrated by the NWI, often prove inadequate due to their inability to accurately delineate these wetland features. To address this critical gap, we utilized higher resolution Sentinel-2 imagery and centered our methodology on a multifaceted approach and employed a diverse set of variables to map the spatial distribution of these wetlands.

Our findings agreed with the results of prior research studies that have underscored the limitations of the NWI in effectively delineating small wetlands. Du et al. [41] conducted an assessment of the NWI program and revealed the challenges faced in achieving precision in wetland mapping, particularly when dealing with the smaller wetland patches – attributable to the NWI adoption of a relatively larger mapping unit, approximately 0.20 hectares. Similarly, Chignell et al. [42] expressed the limitations of the NWI, especially when applied in remote and forested terrains and advocated for the integration of supplementary methodologies to improve the mapping of wetlands. These studies support the compelling necessity for alternative methodologies that could overcome the limitations inherent in traditional wetland mapping approaches.

Numerous studies have documented the challenges in utilizing remote sensing technologies for wetland mapping, particularly in environments characterized by intricate landscapes or dense vegetation cover. Sánchez-Espinosa and Schröder [43] reported difficulties when mapping wetlands using Landsat imagery, a dilemma mostly related to mixed land cover types and spectral confusion. Further, Zhang et al. [32] identified restrictions in the classification of wetlands with Sentinel-2 data within regions categorized by dense vegetation and spectral signatures that exhibit significant overlap. These disagreements in findings underscore the importance of carefully considering the unique characteristics of the study area and the primary challenges associated with remote sensing methods in wetland mapping.

In our study, we overcome the limitations posed by the NWI by utilizing a higher spatial resolution Sentinel-2 imagery, thus enabling a more accurate mapping of small wetlands. This approach provides invaluable insights into the spatial distribution of these ecologically pivotal ecosystems, as represented by the BWG area. However, it is important to acknowledge the limitations of our models, such as potential challenges in classifying complex landscapes and the need for further research to integrate additional data sources into the algorithm.

4.3. Importance of Input Variables

The consistent patterns observed across the spectrum of variables used in the ML models – various spectral bands, spectral indices and image texture attributes – have underscored the importance of specific variables in the accurate delineation of wetlands. Particularly, the B4 and B11 bands, in combination with the NDVI, have emerged as pivotal variables for wetland mapping. The spectral bands B8, B5 and B2 also exhibited importance when discriminating wetlands from other land cover classes. Our results align with the findings by Mahdianpari et al. [22] that emphasized the critical role of SWIR bands and spectral indices in the classification of wetlands. Further, Judah and Hu [37] emphasized the significance of NDVI, particularly when derived from images during the spring months, as a critical variable for wetland classification. The absence of image texture as an important variable in the classification process supports the observations presented by Liu et al. [24] when mapping wetlands within the Tibetan Plateau. These consistencies in research results highlight the importance of specific variables that significantly contribute to the efficacy of wetland classification methodologies.

4.4. Implications

Our research results have considerable implications for conservation, operation and scientific exploration of small-sized wetlands. The accuracy achieved in mapping these wetlands through the utilization of advanced ML techniques and remote sensing data offers critical insights regarding their spatial distribution, characteristics and their ecological significance.

Wetland mapping is a crucial component of conservation and management initiatives, particularly when addressing small-scale wetlands due to their ecological importance as habitats for a diverse array of plant and animal species [44]. The mapping of these wetlands not only facilitates the implementation of targeted conservation strategies, including habitat protection and restoration planning but also ensures the preservation of biodiversity and the continuity of essential ecosystem functions [45]. Furthermore, the mapping of small-sized wetlands proves essential for the comprehensive evaluation and quantification of their contributions to ecosystem services. These ecosystems play an active role in vital processes such as water purification, flood regulation and carbon sequestration [46]. The precision achieved through accurate wetland mapping streamlines the assessment of their spatial distribution, facilitating the precise quantification of their roles in delivering ecosystem services. This, consecutively, encourages decision-making that is more informed and discerning.

This research could provide insights for BWG wetlands habitat assessment and restoration initiatives, particularly given the key role of our study area as a critical habitat for numerous rare and endangered flora and fauna species. The precise mapping of wetlands not only facilitates the identification of habitat areas of utmost importance but also helps in the identification of tailored restoration efforts. These efforts bolster hydrological connectivity and thereby advancing the cause of biodiversity conservation and the rejuvenation of ecosystems [47].

Efforts to monitor of small-sized wetlands could greatly benefit from accurate mapping data [45]. Researchers could utilize these mapped wetlands as study sites to investigate and assess environmental impacts, and monitor changes over time. The availability of reliable wetland maps enhances the efficiency and accuracy of field data collection, optimizing resource allocation and permitting a more focused investigation into specific wetland areas of interest [22,37]. Finally, our results underscore the importance of wetland mapping in supporting policy development and land-use planning. Including accurate wetland information in land use plans and zoning regulations ensures the protection of these ecologically sensitive areas and minimizes potential conflicts between wetland conservation and development activities. Reliable wetland maps aid in assessing the ecological implications of proposed land use changes and support sustainable land management practices.

5. Conclusions

This research highlights the importance and effectiveness of utilizing advanced machine learning techniques and remote sensing data for mapping small-sized wetlands. The findings demonstrated that machine learning classifiers, such as RF and GTB, can accurately classify and map small-sized wetlands. These classifiers outperformed other methods, such as SVM, CART, KNN and NB, in terms of overall accuracy and agreement with reference classifications.

Accurately mapping small-sized wetlands has benefits for wetland conservation, management and research. Apart from enabling targeted conservation efforts, climate change adaptation strategies and habitat assessment and restoration, it is essential for scientific research and informed policy support. Our results and methods could contribute to mapping small-sized wetlands, and provide necessary maps for wetland preservation. The application of advanced machine learning techniques using GEE in small-sized wetland mapping creates ideas for further research and application in other geographic regions and ecosystems. Overall, this study emphasized the significance of accurate wetland mapping using machine learning approaches. It serves as a foundation for practical applications aimed at improving wetland management and promoting their long-term sustainability.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The study was also supported by NIFA/USDA through Central State University Evans-Allen Research Program Fund Number NI201445XXXXG018-0001.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. Finlayson CM, Davidson NC, Spiers AG, et al. (1999) Global wetland inventory – current status and future priorities. *Mar Freshwater Res* 50: 717–727. https://doi.org/10.1071/MF99098
2. Kotze DC, Ellery WN, Macfarlane DM, et al. (2012) A rapid assessment method for coupling anthropogenic stressors and wetland ecological condition. *Ecol Indic* 13: 284–293. https://doi.org/10.1016/j.ecolind.2011.06.023
3. Dahl T, Allord G Technical Aspects of Wetlands. History of Wetlands in the Conterminous United States. Available from: https://water.usgs.gov/nwsum/WSP2425/history.html.
4. Darrah SE, Shennan-Farpón Y, Loh J, et al. (2019) Improvements to the Wetland Extent Trends (WET) index as a tool for monitoring natural and human-made wetlands. *Ecol Indic* 99: 294–298. https://doi.org/10.1016/j.ecolind.2018.12.032
5. Davey Resource Group (2006) GIS Wetlands Inventory and Restoration Assessment, Cuyahoga County, Ohio, Cuyahoga Soil and Water Conservation District.
6. van der Kamp G, Hayashi M (1998) The Groundwater Recharge Function of Small Wetlands in the Semi-Arid Northern Prairies. *Great Plains Res* 8: 39–56.
7. Tang Z, Li Y, Gu Y, et al. (2016) Assessing Nebraska playa wetland inundation status during 1985–2015 using Landsat data and Google Earth Engine. *Environ Monit Assess* 188: 654. https://doi.org/10.1007/s10661-016-5664-x
8. Gibbs JP (1993) Importance of small wetlands for the persistence of local populations of wetland-associated animals. *Wetlands* 13: 25–31. https://doi.org/10.1007/BF03160862
9. Wang W, Sun M, Li Y, et al. (2022) Multi-Level Comprehensive Assessment of Constructed Wetland Ecosystem Health: A Case Study of Cuihu Wetland in Beijing, China. *Sustainability* 14: 13439. https://doi.org/10.3390/su142013439
10. Szantoi Z, Escobedo FJ, Abd-Elrahman A, et al. (2015) Classifying spatially heterogeneous wetland communities using machine learning algorithms and spectral and textural features. *Environ Monit Assess* 187: 262. https://doi.org/10.1007/s10661-015-4426-5
11. Chatziantoniou A, Psomiadis E, Petropoulos GP (2017) Co-Orbital Sentinel 1 and 2 for LULC Mapping with Emphasis on Wetlands in a Mediterranean Setting Based on Machine Learning. *Remote Sens* 9: 1259. https://doi.org/10.3390/rs9121259
12. Wei C, Guo B, Fan Y, et al. (2022) The Change Pattern and Its Dominant Driving Factors of Wetlands in the Yellow River Delta Based on Sentinel-2 Images. *Remote Sens* 14: 4388. https://doi.org/10.3390/rs14174388
13. Malinowski R, Lewiński S, Rybicki M, et al. (2020) Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery. *Remote Sens* 12: 3523. https://doi.org/10.3390/rs12213523
14. Cai Y, Lin H, Zhang M (2019) Mapping paddy rice by the object-based random forest method using time series Sentinel-1/Sentinel-2 data. *Adv Space Res* 64: 2233–2244. https://doi.org/10.1016/j.asr.2019.08.042
15. Jamali A, Mahdianpari M, Brisco B, et al. (2021) Deep Forest classifier for wetland mapping using the combination of Sentinel-1 and Sentinel-2 data. *GIScience Remote Sens* 58: 1072–1089. https://doi.org/10.1080/15481603.2021.1965399

16. Millard K, Richardson M (2013) Wetland mapping with LiDAR derivatives, SAR polarimetric decompositions, and LiDAR–SAR fusion using a random forest classifier. *Can J Remote Sens* 39: 290–307. https://doi.org/10.5589/m13-038
17. Onojeghuo AO, Onojeghuo AR, Cotton M, et al. (2021) Wetland mapping with multi-temporal sentinel-1 & -2 imagery (2017–2020) and LiDAR data in the grassland natural region of alberta. *GIScience Remote Sens* 58: 999–1021. https://doi.org/10.1080/15481603.2021.1952541
18. Pham H-T, Nguyen HQ, Le KP, et al. (2023) Automated Mapping of Wetland Ecosystems: A Study Using Google Earth Engine and Machine Learning for Lotus Mapping in Central Vietnam. *Water* 15: 854. https://doi.org/10.3390/w15050854
19. Bhatnagar S, Gill L, Regan S, et al. (2020) Mapping vegetation communities inside wetlands using Sentinel-2 imagery in Ireland. *Int J Appl Earth Obs Geoinformation* 88: 102083. https://doi.org/10.1016/j.jag.2020.102083
20. Ruiz LFC, Guasselli LA, Simioni JPD, et al. (2021) Object-based classification of vegetation species in a subtropical wetland using Sentinel-1 and Sentinel-2A images. *Sci Remote Sens* 3: 100017. https://doi.org/10.1016/j.srs.2021.100017
21. Kaplan G, Avdan U (2019) Evaluating Sentinel-2 Red-Edge Bands for Wetland Classification. *Proceedings* 18: 12. https://doi.org/10.3390/ECRS-3-06184
22. Mahdianpari M, Jafarzadeh H, Granger JE, et al. (2020) A large-scale change monitoring of wetlands using time series Landsat imagery on Google Earth Engine: a case study in Newfoundland. *GIScience Remote Sens* 57: 1102–1124. https://doi.org/10.1080/15481603.2020.1846948
23. Waleed M, Sajjad M, Shazil MS, et al. (2023) Machine learning-based spatial-temporal assessment and change transition analysis of wetlands: An application of Google Earth Engine in Sylhet, Bangladesh (1985–2022). *Ecol Indic* 75: 102075. https://doi.org/10.1016/j.ecoinf.2023.102075
24. Liu Q, Zhang Y, Liu L, et al. (2021) A novel Landsat-based automated mapping of marsh wetland in the headwaters of the Brahmaputra, Ganges and Indus Rivers, southwestern Tibetan Plateau. *Int J Appl Earth Obs Geoinformation* 103: 102481. https://doi.org/10.1016/j.jag.2021.102481
25. Wu N, Crusiol L, Liu G, et al. (2023) Comparing Machine Learning Algorithms for Pixel/Object-Based Classifications of Semi-Arid Grassland in Northern China Using Multisource Medium Resolution Imageries. *Remote Sens* 15: 750. https://doi.org/10.3390/rs15030750
26. Gemechu GF, Rui X, Lu H (2022) Wetland Change Mapping Using Machine Learning Algorithms, and Their Link with Climate Variation and Economic Growth: A Case Study of Guangling County, China. *Sustainability* 14: 439. https://doi.org/10.3390/su14010439
27. Ghorbanian A, Zaghian S, Asiyabi RM, et al. (2021) Mangrove Ecosystem Mapping Using Sentinel-1 and Sentinel-2 Satellite Images and Random Forest Algorithm in Google Earth Engine. *Remote Sens*13: 2565. https://doi.org/10.3390/rs13132565
28. Hemati MA, Hasanlou M, Mahdianpari M, et al. (2021) Wetland Mapping of Northern Provinces of Iran Using Sentinel-1 and Sentinel-2 in Google Earth Engine, *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 96–99. https://doi.org/10.1109/IGARSS47720.2021.9554984
29. Gxokwe S, Dube T, Mazvimavi D (2022) Leveraging Google Earth Engine platform to characterize and map small seasonal wetlands in the semi-arid environments of South Africa. *Sci. Total Environ* 803: 150139. https://doi.org/10.1016/j.scitotenv.2021.150139

30. Tamiminia H, Salehi B, Mahdianpari M, et al. (2020) Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS J Photogramm Remote Sens* 164: 152–170. https://doi.org/10.1016/j.isprsjprs.2020.04.001
31. Pal M (2005) Random forest classifier for remote sensing classification. *Int J Remote Sens* 26: 217–222. https://doi.org/10.1080/01431160412331269698
32. Zhang L, Hu Q, Tang Z (2022) Assessing the contemporary status of Nebraska's eastern saline wetlands by using a machine learning algorithm on the Google Earth Engine cloud computing platform. *Environ Monit Assess* 194: 193. https://doi.org/10.1007/s10661-022-09850-8
33. Domingos P, Pazzani M (1997) On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Mach Learn* 29: 103–130. https://doi.org/10.1023/A:1007413511361
34. Hall P, Park BU, Samworth RJ (2008) Choice of neighbor order in nearest-neighbor classification. *Ann Stat* 36: 2135–2152. https://doi.org/10.1214/07-AOS537
35. Boser B, Guyon I, Vapnik V (1992) A Training Algorithm for Optimal. Margin Classifiers. *COLT*, 144–152. https://doi.org/10.1145/130385.130401
36. Fekri E, Latifi H, Amani M, et al. (2021) A Training Sample Migration Method for Wetland Mapping and Monitoring Using Sentinel Data in Google Earth Engine. *Remote Sens* 13: 4169. https://doi.org/10.3390/rs13204169
37. Judah A, Hu B (2022) The Integration of Multi-Source Remotely Sensed Data with Hierarchically Based Classification Approaches in Support of the Classification of Wetlands. *Can J Remote Sens* 48: 158–181. https://doi.org/10.1080/07038992.2021.1967732
38. Zhang M, Lin H (2022) Wetland classification using parcel-level ensemble algorithm based on Gaofen-6 multispectral imagery and Sentinel-1 dataset. *J Hydrol* 606: 127462. https://doi.org/10.1016/j.jhydrol.2022.127462
39. Amani M, Salehi B, Mahdavi S, et al. (2017) Wetland classification in Newfoundland and Labrador using multi-source SAR and optical data integration. *GIScience Remote Sens* 54: 779–796. https://doi.org/10.1080/15481603.2017.1331510
40. Xing H, Niu J, Feng Y, et al. (2023) A coastal wetlands mapping approach of Yellow River Delta with a hierarchical classification and optimal feature selection framework. *CATENA* 223: 106897. https://doi.org/10.1016/j.catena.2022.106897
41. Du L, McCarty GW, Zhang X, et al. (2020) Mapping Forested Wetland Inundation in the Delmarva Peninsula, USA Using Deep Convolutional Neural Networks. *Remote Sens* 12: 644. https://doi.org/10.3390/rs12040644
42. Chignell SM, Luizza MW, Skach S, et al. (2018) An integrative modeling approach to mapping wetlands and riparian areas in a heterogeneous Rocky Mountain watershed. *Remote Sens Ecol Conserv* 4: 150–165. https://doi.org/10.1002/rse2.63
43. Sánchez-Espinosa A, Schröder C (2019) Land use and land cover mapping in wetlands one step closer to the ground: Sentinel-2 versus landsat 8. *J Environ Manage* 247: 484–498. https://doi.org/10.1016/j.jenvman.2019.06.084
44. Sebastián-González E, Green AJ (2014) Habitat Use by Waterbirds in Relation to Pond Size, Water Depth, and Isolation: Lessons from a Restoration in Southern Spain. *Restor Ecol* 22: 311–318. https://doi.org/10.1111/rec.12078
45. Gitau P, Ndiritu G, Gichuki N (2019) Ecological, recreational and educational potential of a small artificial wetland in an urban environment. *Afr J Aquat Sci* 44: 329–338. https://doi.org/10.2989/16085914.2019.1663721

46. Jie Y, Zhao Y (2021) Trends in Research on Wetland Restoration based on Web of Science Database. *Ecol Environ* 30: 1541–1548. https://doi.org/10.16258/j.cnki.1674-5906.2021.07.023
47. Rebelo AJ, Scheunders P, Esler KJ, et al. (2017) Detecting, mapping and classifying wetland fragments at a landscape scale. *Remote Sens Appl Soc Environ* 8: 212–223. https://doi.org/10.1016/j.rsase.2017.09.005