



Deep Learning for Automatic Extraction of Water Bodies Using Satellite Imagery

Reham Gharbia^{1,2} 

Received: 26 October 2022 / Accepted: 18 April 2023 / Published online: 11 June 2023
© The Author(s) 2023

Abstract

The study introduces an automated approach for extracting water bodies from satellite images using the Faster R-CNN algorithm. The approach was tested on two datasets consisting of water body images collected from Sentinel-2 and Landsat-8 (OLI) satellite images, totaling over 3500 images. The results showed that the proposed approach achieved an accuracy of 98.7% and 96.1% for the two datasets, respectively. This is significantly higher than the accuracy achieved by the convolutional neural network (CNN) approach, which achieved 96% and 80% for the two datasets, respectively. These findings highlight the effectiveness of the proposed approach in accurately mapping water bodies from satellite imagery. Additionally, the Sentinel-2 dataset performed better than the Landsat dataset in both the Faster R-CNN and CNN approaches for water body extraction.

Keywords Convolutional neural network · Faster R_CNN · Water bodies · Water bodies extraction · Water mapping

Introduction

Water bodies extraction is critical in hydrology and water resources management. Accurate and up-to-date information on the location, size, and distribution of water bodies such as lakes, rivers, and wetlands can provide insights into the quantity and quality of available water resources. This information is important for the management of water resources, including water allocation, flood management, and water quality monitoring. Water body extraction plays an essential role in environmental studies, including ecology, biology, and geology. Accurate data on water bodies can provide insights into environmental conditions, habitat suitability, and species distribution. This information is critical for identifying conservation priorities, understanding ecological processes, and assessing the impacts of human activities on the environment (Lacaux et al., 2007;

Viala, 2008). It is also important for climate change studies. Water bodies, particularly lakes, and reservoirs, play a critical role in the global carbon cycle and the exchange of greenhouse gases between the atmosphere and water bodies. Extracting accurate data on the size and distribution of water bodies is important for understanding the role of water bodies in the global carbon cycle, predicting the impacts of climate change on water resources, and designing effective adaptation strategies (Enan, 2021). In addition, water body mapping can be crucial in disaster management and emergency response, such as during natural disasters like floods or hurricanes. Accurate water body extraction can help identify areas that are at risk of flooding and provide information on potential evacuation routes (Wang, 2021; Qin, 2019).

Satellite imagery is a powerful tool for mapping water bodies, and its use has been demonstrated in various studies across the globe (Frazier & Page, 2000; Santoro et al., 2015). The availability of high-resolution satellite images and advanced image processing techniques can provide accurate and detailed information on water body dynamics, distribution, and quality (Gharbia et al., 2018; Verpoorter et al., 2014). Das et al. (2020) utilized satellite imagery to map water bodies within the Brahmaputra River basin in India. The study found that high-resolution satellite images

✉ Reham Gharbia
drrehamgharbia@gmail.com
<http://www.egyptscience.net>

¹ Nuclear Materials Authority,
P.O.B.: 550 Maadi, Cairo 11728, Egypt

² Member of the Scientific Research Group in Egypt, Cairo,
Egypt

can be used to accurately map water bodies and assess their dynamics. Sentinel-2 satellite images were used to map water bodies in the Three Gorges Reservoir, China. The study found that Sentinel-2 imagery can provide accurate and detailed information on water body dynamics (Bie, 2020). Several methods have been suggested for recognizing water bodies by utilizing Landsat bands in various water indices (Du et al., 2012; Rokni et al., 2014). Meanwhile, some investigations have employed high-resolution datasets such as Indian Remote Sensing data of RS2-L4 to revise the glacier lake inventory in the Indus River basin. This was accomplished by manually outlining the borders of the lakes and visually interpreting them (Gupta et al., 2022).

Since the 1990s, numerous supervised and unsupervised methods have been utilized to identify water bodies using Landsat Enhanced Thematic Mapper data; unsupervised image classification was found to provide the highest accuracy but was limited at lake edges (Elsahabi et al., 2016; Guo et al., 2017; Yan, 2018; Yang & Chen, 2017). Several water indexes techniques have been proposed, such as the normalized difference water index (NDWI) based on Landsat which has been used to analyze seasonal variations over a 15-year period (2001–2015) in the Telibandha Lake Catchment (Kumari et al., 2022). However, shadows in built-up areas can interfere with NDWI's performance. To address this issue, Xu proposed an adjusted standardized contrast water index (MNDWI) which uses shortwave infrared (SWIR) band instead of a NIR band (Xu, 2006). But this model is unable to distinguish between water bodies and shadows (Huang et al., 2015). The limited spectral resolution of high-resolution remote-sensing images generally makes it challenging to implement improvements in the water index (Chen et al., 2018).

In general, unsupervised classification techniques involve k-means clustering and ISODATA clustering approaches, while machine-learning techniques typically include neural networks and support vector machines (Katz, 2016). One limitation of machine learning in water body extraction from remote sensing is the reliance on accurately labeled training data. This can be challenging to obtain, as manually labeling large amounts of data can be time-consuming and resource intensive. Additionally, variations in water body types, sizes, and environmental conditions can make it difficult to create a model that can be trained to perform effectively across a variety of contexts. Another limitation is the complexity of the algorithms used, which can make it difficult to interpret and explain the results. Furthermore, the accuracy of the model is highly dependent on the quality and resolution of the input remote-sensing data, which can limit its applicability in certain scenarios. Traditional machine-learning techniques typically involve the manual selection of features

from the imagery, such as texture, color, and shape. These features are then used to train a model, such as a support vector machine, to classify pixels as water or non-water. Despite its effectiveness, the semantic segmentation approach can be time-consuming and requires domain expertise to identify and select the appropriate features, as noted by Mao et al. (2019). On the other hand, deep learning approaches utilize convolutional neural networks (CNNs) to discover more distinctive spatial-spectral characteristics, leading to improved detection and identification of water bodies (Isikdogan et al., 2017). Convolutional neural networks (CNNs) have been widely used in recent years to recognize objects such as water bodies from high-resolution satellite imagery (Chen et al., 2020; Dong et al., 2019; Gharbia et al., 2021; Wang et al., 2020; Zhang et al., 2021). Compared to ratio-based indexes for extracting water bodies, CNNs offer significant multi-scale feature representation capabilities and can effectively extract both low-level location information and high-level semantic information. When it comes to detecting continuous surface objects, such as water bodies and highways, semantic segmentation is generally considered the most appropriate approach. Models commonly used for this purpose include DeepLab (Chen et al., 2017), UNet (Ronneberger et al., 2015), PSPNet (Zhou et al., 2017), and SegNet (Badrinarayanan et al., 2017). It has been demonstrated that deep learning methods are generally more effective than shallow classification methods, such as SVM, as noted by LeCun et al. (2015). Cheng et al. (2017) used CNN (convolutional neural network) instead of manually engineered features to segment water bodies in their study. On the other hand, Lin et al. (2017) employed a fully convolutional network (FCN) to incorporate multi-scale information into their research. They designed a multilayer deconvolutional network to address the scale challenge (Noyola-Medrano & Martínez-Sías, 2017). Autoencoders have been used to extract high-level feature maps from high-resolution images (Miao et al., 2018), while the ResNet model has been employed to identify global water reservoirs. Additionally, related regions with convolutional neural networks (R-CNN) models have been widely used in the field of object detection (Fang et al., 2019).

Faster R-CNN is a type of object detection model that uses a CNN as a backbone network. It combines the region proposal network (RPN) with a detection network to perform both object localization and classification. Faster R-CNN is known for its speed and accuracy and has been used in various applications such as autonomous vehicles, surveillance, and robotics. While CNNs are primarily used for image classification, Faster R-CNNs are used for object detection tasks and have the added advantage of being faster and more accurate than other object detection models (Lee et al., 2016).

Faster R-CNN has been shown to significantly improve overall performance by incorporating the region proposal network (RPN), particularly in terms of detection speed. The Faster R-CNN is primarily designed for object detection (Ren et al., 2015), and there are several research studies that have explored using Faster R-CNN for image classification. Researchers have modified the Faster R-CNN algorithm in these studies to work with fixed image regions rather than region proposals. The Faster R-CNN algorithm was used to improve both object detection and image classification. In their study, Wang et al. (2019) made modifications to the RPN (region proposal network) by creating a fixed set of anchor boxes, as opposed to region proposals. These anchor boxes were utilized to generate predetermined image regions for the purposes of classification. The proposed algorithm is evaluated on several standard image classification benchmarks, achieving state-of-the-art results.

This paper proposes an automated approach for extracting water bodies from satellite imagery using Faster R-CNN. This method was applied to extract water bodies from Sentinel-2 and Landsat-8 (OLI) data for comparison. The paper is arranged as follows: Second section outlines the materials and methods, the Third section presents the results and discussion, and the fourth section provides conclusions and future work.

Materials and Methods

Study Area

Regarding the importance of covering various types and sizes of water bodies to assess the proposed approach, Egypt is the suitable choice for this study, because it includes abundant and varied bodies of water and various urban environments. According to El-Rawy et al. (2020), Egypt is situated in Northeast Africa and has a land area of approximately one million square kilometers. Egypt has both the Mediterranean and the Red Sea as part of its borders. This country borders, Libya and Sudan to the west and south Fig. 1. Egypt's water resources include the Nile River, rainfall, and flash flooding. Fresh water in Egypt is derived from the Nile River Nile, which provides about 97% of the country's supply. Egypt has riverine and coastal rivers (Yu et al., 2019):

The Qattara Depression is the northern inner basin covering a total area of 520,881 km², approximately 52%. The area of the Nile Basin is 326,751 km², about 33% of the country's total area. The Mediterranean Coast Basin is 65,568 km², about 6% of its total area and the Red Sea Basin is an area of roughly 438,000 km².

The Proposed Approach

The paper suggests using the Faster R-CNN (region-based convolutional neural network) for water bodies extraction, which involves two phases as follows; see Fig. 2:

1. Preprocessing
2. Faster R_CNN for water bodies extraction

The proposed Faster R_CNN to extract the water bodies approach is trained and validated on different satellite datasets, Landsat-8 (OLI) and Sentinel-2 satellite.

Dataset and Preprocessing

Optical satellite systems have been used the most in water body extraction science. The visible and near infrared (VNIR) ranges from 0.4 to 1.3 m, the shortwave infrared (SWIR) ranges from 1.3 to 3.0 m, the thermal infrared (TIR) ranges from 3.0 to 15.0 m, and the long-wavelength infrared (LWIR) ranges by (7–14 m) are both protected by these sensors. The proposed approach was trained and validated on different satellite datasets, Landsat-8 (OLI) and Sentinel-2 satellite. Sentinel-2A is a satellite that carries a multispectral instrument. The MSI instrument has a ten-day repeat interval over a broadband of 290 km. In February 2013, the Landsat-8 satellite launched loads of the Thermal Infrared Sensor and Operational Land Imager (OLI) that observes the Planet's land surface over 185 km in a repeat cycle. Both sensors have a 12-bit dynamic resolution and are rarely saturated on highly reflective surfaces. Landsat-8 OLI and Sentinel-2A MSI have a greater than 97 percent association between their measurements. The MSI Sentinel-2 is equipped with bands of 13 reflective wavelengths: four visible bands with a resolution of 10 m and a near-infrared (NIR) band with a resolution of 10 m, 6 red edges, NIR, and (SWIR) short-wave infrared bands with a resolution of 20 m, and other bands for resolution of 60 m. The Landsat-8 (OLI) owns nine wavelength bands, six of them with 30 m spatial resolution (visible, NIR, SWIR1, SWIR2) destined for land applications.

The images are downloaded individually from the satellite. Several preprocessing processes must be completed or treated effectively before using Sentinel-2 and Landsat-8. More than ten scenes for each Landsat-8 (OLI) and Sentinel-2 satellite cover Egypt, and the scenes have been downloaded from the USGS Earth Explorer and ESA Sentinel-2 Pre-Operations Hub, respectively. A clear non-hazy day, minimizing potential impact from the atmosphere, was set for all images; see the characteristics of the bands in each scene. The dataset was splatted into two datasets.



Fig. 1 Egypt Map

The first dataset is the Sentinel-2 images. The Sentinel-2 Level-1C production, which was provided by geometric and radiometric corrections, contains ortho-rectification and spatial registration. The Level-1C of Sentinel-2 production in the UTM/WGS84 projection is formed in $100\text{ km} \times 100\text{ km}$. Ten scenes of the Sentinel-2 Level-1C image were obtained from 1st July to 1st Oct. 2020. All bands were resampled by using SNAP software and the false-color composite (FCC) image of the Sentinel-2 at 10 m is produced from the green band (3), red band (4), and NIR band (8). The FCC is produced in layer stack in Imagine Erdas software. The satellite image covered a very large area, usually reaching larger than $20,000\text{ m} \times 20,000\text{ m}$. Therefore, it is difficult to process the full image at the same time. The FCC image is used to

prepare the training data by subset it into two classes water and no-water. The water objects are specified by the presence of light black on the FCC (band 8 4 3). Each subset image is $(226 \times 226 \times 3)$ pixels. The number of training data reaches more than 1500 images for the first dataset.

The second dataset is the Landsat-8 (OLI) level-2 product, which is provided through geometric and atmospheric corrections. The Level-2 product that is Landsat-8 (OLI) is in the UTM/WGS84 projection. Ten scenes of Landsat-8 (OLI) images were obtained from 1st July to 1st Oct. 2020. All bands were resampled by using Imagine Erdas software and the false-color composite (FCC) image of the Landsat-8 (OLI) at 30 m is produced from bands 5, 4, and 3 respectively. As the first dataset, the FCC image is used to

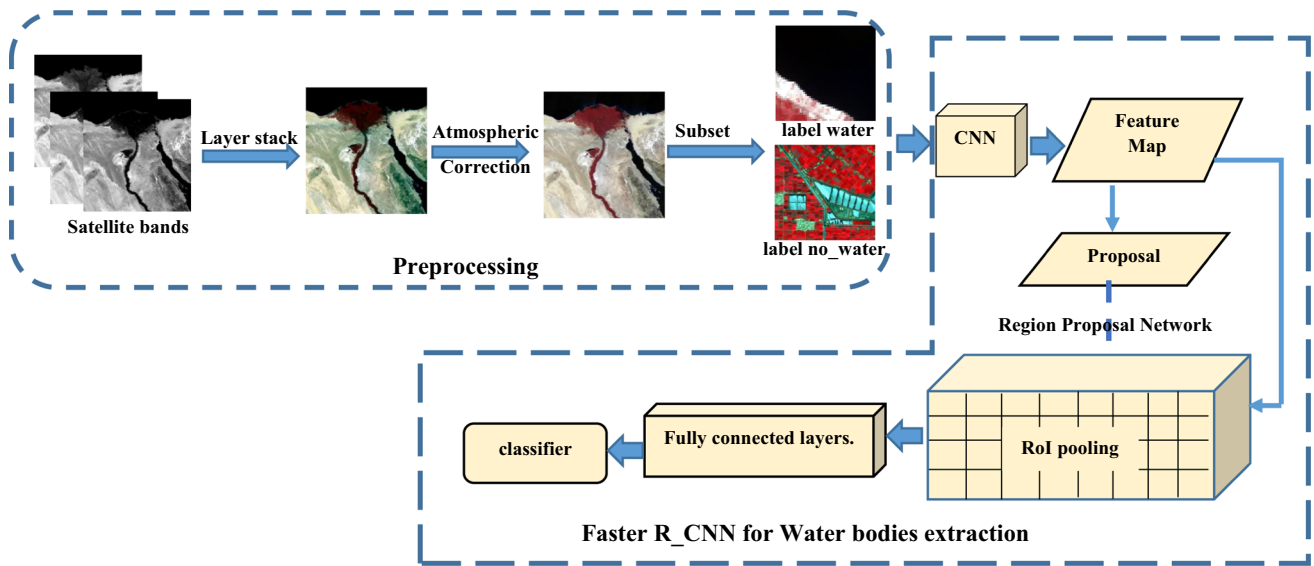


Fig. 2 The framework of the proposed Faster R-CNN for water bodies extraction approach

prepare the training data by subset it into two classes water (black) and no-water (red) based on the features; see Fig. 3. The water objects are specified by the presence of light black on the FCC (band 5 4 3). Each image is (226 × 226 × 3) pixels. The number of training data reaches more than 1500 images for the second dataset.

Faster R-CNN for Water Bodies Extraction

In this section, we provide a brief overview of the fundamental components of Faster R-CNN. For more in-depth technical information, we suggest referring to the original paper (Ren et al., 2015). The RPN involves the use of convolution layers from a pre-trained network, followed by

a 3 × 3 convolutional layer. This effectively reduces the dimensionality of a large spatial window (e.g., 228 × 228) in the input image to a lower-dimensional feature vector at a center stride (e.g., 16). Next, two 1 × 1 convolutional layers are added for both the classification and regression branches of all spatial windows.

The proposed Faster R-CNN approach for water bodies extraction consists of combining a region proposal network (RPN) with a Fast-RCNN to detect and localize water bodies in an image as the following steps:

1. Input label image into the network.
2. Use a convolutional neural network (CNN) to extract features from the image.

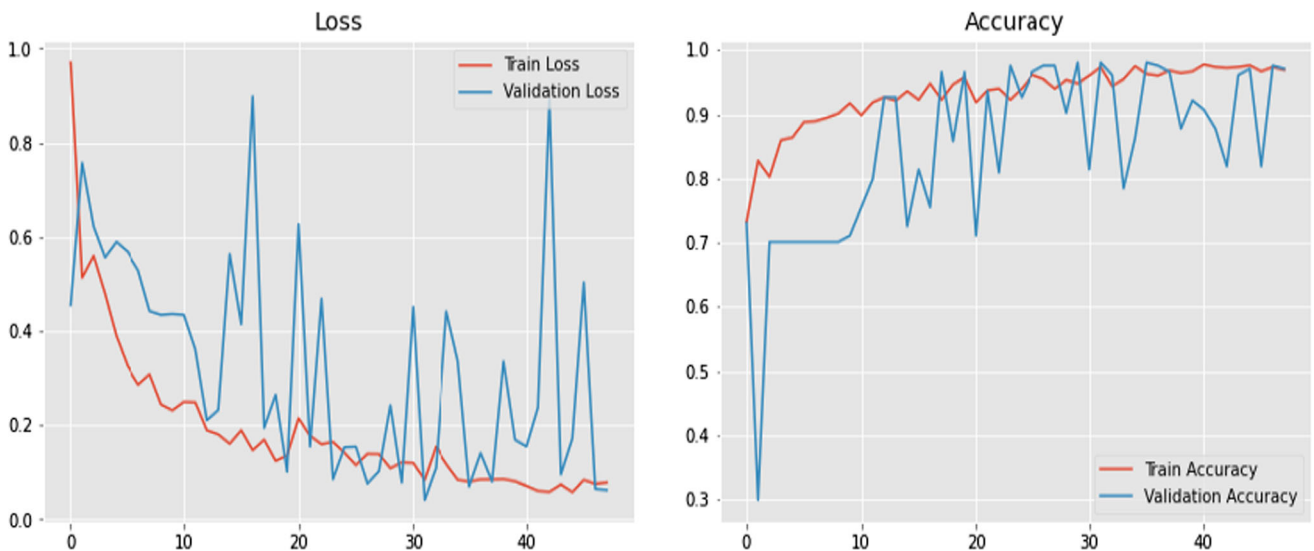


Fig. 3 Loss value and accuracy of the training and validation for the Landsat-8 (OLI) dataset

3. Use a region proposal network (RPN) to generate candidate object regions in the image.
4. Use a region of interest (RoI) pooling layer to extract features from each candidate region.
5. Pass the extracted features through a series of fully connected layers to classify the object and predict its bounding box.
6. Apply the non-maximum suppression (NMS) technique to eliminate redundant detections and produce the ultimate detection outcomes.
7. These steps can be further optimized and fine-tuned through transfer learning, data augmentation, and hyperparameter tuning techniques.

The training is done as the following:

1. Training, validation, and testing: The images and label arrays are randomly shuffled. This allows the images and their corresponding labels to remain linked even after shuffling. The dataset is split into 3 datasets. The training and evaluation datasets are used to train the model, while the testing dataset is used to assess the model on unseen data. Unseen data are applied for simulating real-world prediction, as the model has not seen this data before. The dataset is partitioned into three subsets, with 70% allocated for training, 20% for validation, and 10% for testing. During training, the network is set to run 50 epochs with a set size of 16.
2. The loss and accuracy for the training and validation datasets: A loss function is used to help a machine-learning algorithm to optimize. The loss is based on training and validation, and how well the model is performing in these two test sets. It is the sum of all mistakes made in all training and test sets. A loss estimate is a method that describes how badly or well the model behaves after each cycle of optimization, Figs. 3 and 4 show the loss and accuracy for the training and validation of Landsat-8 (OLI) and Sentinel-2 datasets, respectively.

Evaluation Metrics

There are several metrics that can be used to measure the performance of a deep learning binary classification model. Some commonly used metrics include precision, recall, and overall accuracy (Simonyan & Zisserman, 2014).

Precision

Precision is a performance metric used in binary classification models that measures the accuracy of positive predictions. It is the ratio of true-positive predictions to the total number of positive predictions made by the model. In

other words, precision represents the proportion of positive predictions that are correct. It is a measure of the model's ability to avoid making false-positive predictions, or in other words, its ability to correctly identify true-positive cases. The precision formula is expressed as:

$$\text{Precision} = \frac{\text{Number of True Positives}}{(\text{Number of True Positives} + \text{Number of False Positives})}$$

True positives refer to the count of positive instances that the model correctly predicted, while false positives represent the count of negative instances that the model mistakenly predicted as positive. Precision ranges between 0 and 1, where a higher value indicates better performance. A precision of 1.0 means that all positive predictions made by the model are correct, while a precision of 0.0 means that none of the positive predictions are correct.

Recall

Recall is a performance metric used in binary classification models that measures the ability of the model to correctly identify all positive cases. It is the ratio of true-positive predictions to the total number of actual positive cases in the dataset. In other words, recall represents the proportion of actual positive cases that the model correctly identified as positive. It is a measure of the model's ability to avoid making false-negative predictions, or in other words, its ability to correctly identify true-positive cases. The formula for the recall is:

$$\text{Recall} = \frac{\text{True Positives}}{(\text{True Positives} + \text{False Negatives})}$$

where true positives are the number of positive cases that the model correctly predicted, and false negatives are the number of positive cases that the model incorrectly predicted as negative. Recall ranges between 0 and 1, where a higher value indicates better performance. A recall of 1.0 means that the model correctly identified all positive cases in the dataset, while a recall of 0.0 means that the model did not identify any positive cases.

F1 Score

The F1 score is a commonly used metric for evaluating the performance of binary classification models. It is a measure of the balance between precision and recall, which are two important evaluation metrics in machine learning. The F1 score is the harmonic mean of precision and recall, and it provides a single score that combines these two metrics. The F1 score formula combines precision and recall into a single metric that varies from 0 to 1, where a higher value indicates better performance. It is calculated using the formula:

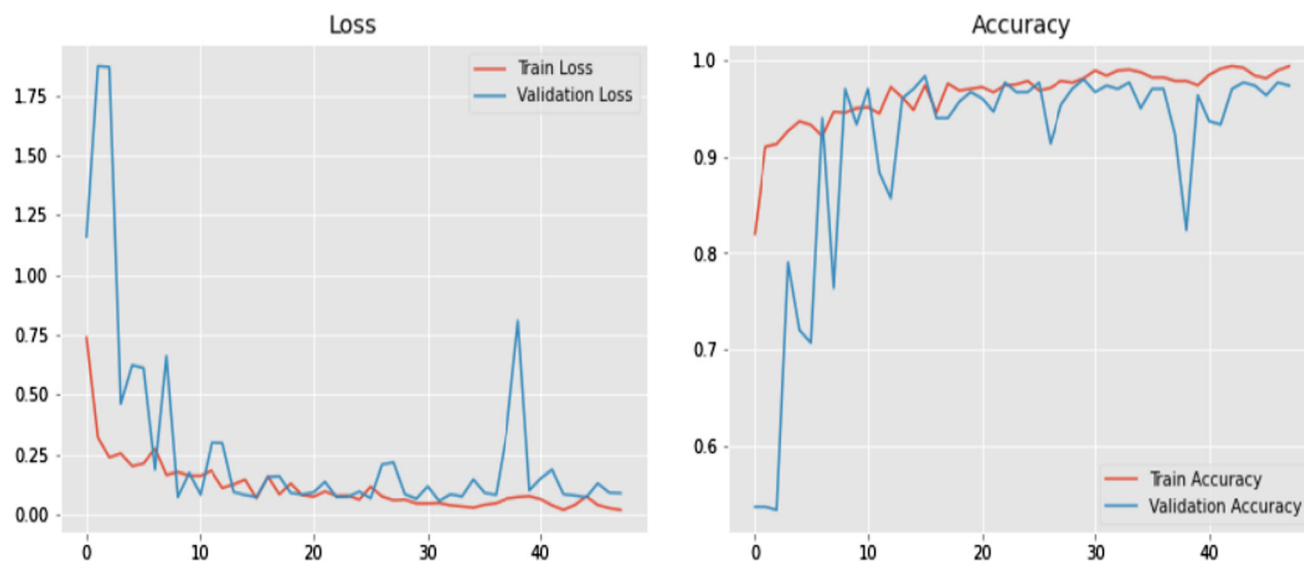


Fig. 4 Loss value and accuracy of the training and validation for the Sentinel-2 dataset

$$F1 \text{ score} = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

Overall Accuracy

Overall accuracy is a performance metric used to evaluate the performance of classification models. It measures the proportion of correct predictions made by the model across all classes in the dataset (Uijlings et al., 2013). The formula for overall accuracy is:

$$\text{Overall accuracy} = (\text{Number of Correct Predictions}) / (\text{Total Number of Predictions})$$

where the number of correct predictions is the sum of true positives and true negatives, and the total number of predictions is the sum of true positives, false positives, true negatives, and false negatives. Overall accuracy varies from 0 to 1, where a higher value indicates better performance. A perfect model would have an accuracy of 1.0, meaning that it correctly predicted all cases in the dataset. While overall accuracy can be a useful metric, it may not always provide a complete picture of a model's performance.

Results and Discussion

Satellite image classification was executed to characterize the classes of (water and no-water), based on the Faster R_CNN described in Fig. 2; also. It uses the Faster R_CNN approach to compare Sentinel-2 and Landsat-8 (OLI) datasets. To show which are more perfect for automatic extraction of the water body. The water class encompasses

significant and distinct features of surface water, such as the sea, streams, lakes, canals, or rivers the training data were assigned to each land category (urban, rivers, lakes, canals, fields, and forests) by CORINE land cover terminology (Tao et al., 2016). The class designated for each dataset was confirmed using Google Earth. The dataset consists of a subset of images extracted from Sentinel-2 and Landsat-8 (OLI) satellite imagery collected over Egypt country. It includes approximately 1500 images for each dataset. The training images are labeled with either a “water” or “no-water” classification. The Landsat-8 (OLI) dataset has 1525 total images, 765 images for the no-water class, and 760 images for the water class; see Fig. 5.

The Sentinel-2 dataset has 1504 total images, 754 images for the no-water class, and 750 images for the water class; see Fig. 6.

In our model, the data are balanced in two datasets (Landsat-8 (OLI) and Sentinel-2) in different classes (water and no_water); see Figs. 5 and 6. Balanced data refers to a situation where the classes or categories in a dataset are equally represented. In machine learning, balanced data are important because it helps ensure that the model is trained on an unbiased representation of the underlying data. When the dataset is balanced, the model is trained equally on all classes, which helps to prevent it from becoming biased towards any particular class. This is important because biased models can lead to inaccurate predictions and decisions. In addition, balanced data can improve the quality of the model's training by reducing the risk of overfitting; see Figs. 3 and 4. Overfitting transpires when the model tends to memorize the data in the training set rather than understanding the underlying patterns present in the data, which can result in poor performance on new,

Fig. 5 The Landsat-8 (OLI) dataset distribution per class

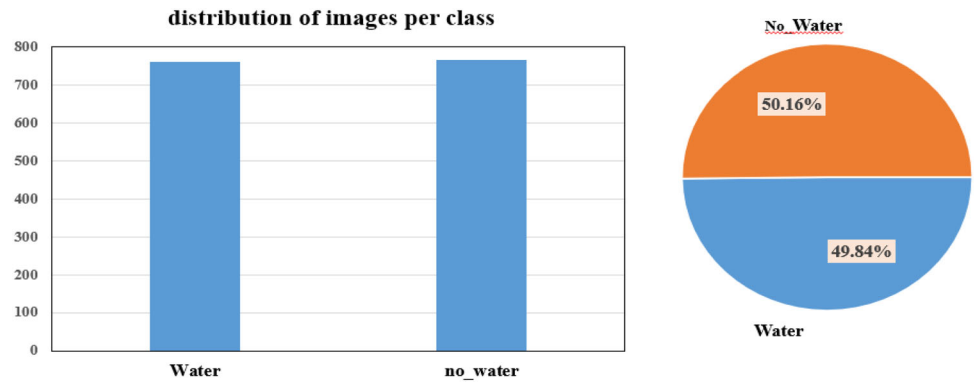
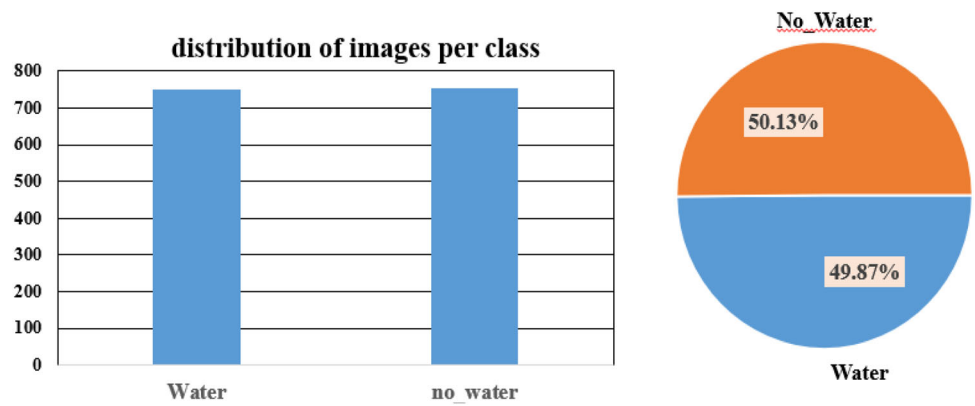


Fig. 6 The Sentinel-2 dataset distribution per class

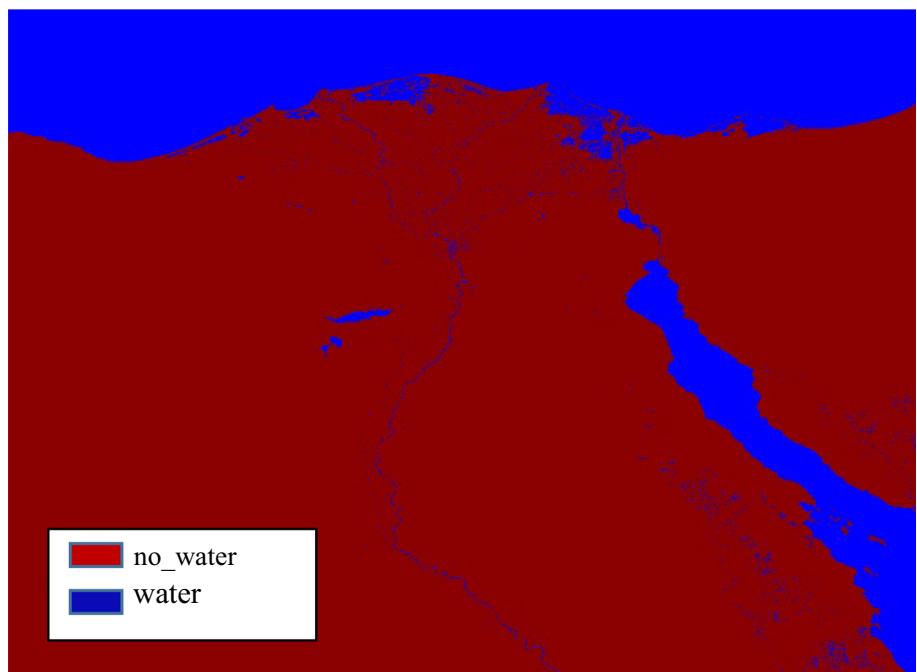


unseen data. With balanced data, the model is less likely to overfit to any class, which can help to improve its generalization performance. Figure 7 demonstrates the

application of the Faster R-CNN approach on a subset of Sentinel-2 data of Egypt.

Recent research in this domain has focused on applying deep learning methods to improve the precision of water

Fig. 7 The sentinel-2 classified image based on the Faster R-CNN



body extraction from satellite images. One such study by Yan et al. (2022) proposed a U-Net architecture for extracting water bodies from Sentinel-2 images, with an F1 score of 0.90 and a relative error of 3%. Similarly, Singh et al. (2021) combined a CNN with a fully connected conditional random field model to extract water bodies from Landsat-8 images, achieving an overall accuracy of 94.5%. Additionally, Yu et al. (2022) proposed a hierarchical attentive high-resolution network that achieved high accuracy with an average precision of 98.44% and an average recall of 97.84%. These recent works demonstrate the potential of deep learning methods in the field of water body extraction from satellite imagery and highlight the importance of exploring new architectures and algorithms to further enhance the accuracy of water body extraction.

To compare the proposed Faster R-CNN approach for water extraction with other existing approaches, such as CNN-based approaches, it is common practice to evaluate the detection performance using a common dataset and consistent evaluation criteria. Metrics such as precision, recall, F1 score, and overall accuracy can be used for comparison. It is crucial to ensure a fair comparison by using the same evaluation criteria, dataset, and preprocessing steps. In addition, the computational resources required to train and run the models, as well as the models' complexity and ability to generalize to different environments, should also be considered. Faster R-CNN is a two-stage architecture, while CNN is a single-stage architecture. Faster R-CNN uses a region proposal network (RPN) to generate regions of interest (RoIs) and then classifies and refines the RoIs using a separate detection network. In contrast, CNN directly classifies water bodies and their locations in a single pass. The Faster R-CNN for water bodies extraction approach is designed specifically for water body detection, while CNN is a more general-purpose model used for a variety of computer vision tasks, including image classification and segmentation. Faster R-CNN achieves higher accuracy in object detection tasks compared to CNN. This is because Faster R-CNN uses a more complex architecture that can accurately localize objects and reduce false positives.

Tables 1 and 2 present a comparison between the Faster R-CNN and CNN approaches implemented on two datasets: Landsat-8 (OLI) and Sentinel-2. The results show that the Faster R-CNN approach for water bodies extraction achieved higher accuracy than the CNN approach based on different metrics, such as precision, recall, F1 score, and overall accuracy on the Landsat-8 (OLI) dataset. On the other hand, the Faster R-CNN approach achieved comparable results in the Sentinel-2 and Landsat-8 (OLI) datasets. The sentinel-2 dataset achieved result better than Landsat-8 (OLI) datasets based on CNN and Faster R-CNN.

Table 1 The performance of the Landsat-8 (OLI) dataset

	F1_score	Precision	Recall	Overall accuracy
F_RCNN	0.963	0.975	0.952	0.961
CNN	0.876	0.808	0.985	0.805

Table 2 The performance of the Sentinel-2 dataset

	F1_score	Precision	Recall	Overall accuracy
F_RCNN	0.967	0.974	0.961	0.987
CNN	0.824	1.000	0.700	0.964

Bold indicates to clarify the results of the proposed approach in comparison with the results of CNN approach

Conclusions and Future Work

This study introduces a Faster R-CNN approach for water extraction from satellite imagery on two datasets, Sentinel-2 and Landsat-8 (OLI), containing over 1500 images each. The training dataset was labeled with a “water” or “no-water” classification. The Faster R-CNN approach outperformed the CNN approach in terms of precision, recall, F1_score, and overall accuracy on both datasets, with the Sentinel-2 dataset performing better than Landsat-8 (OLI) for both approaches. Comparative tests demonstrated the practical feasibility and advanced superiority of the Faster R-CNN approach in water body extraction tasks. Although the Faster R-CNN model proposed in this study has shown promise in water body extraction from remote sensing images, there are still limitations that affect its accuracy in localizing and segmenting water bodies. To address these limitations, future research may investigate advanced techniques such as super-resolution to improve the quality of extraction for small and narrow water bodies, spatial context characterization strategies to enhance the extraction of occluded water body regions, and weakly supervised or few-shot learning architectures to reduce the need for large amounts of training data.

Acknowledgment The authors would like to extend their sincere gratitude to the editors and anonymous reviewers for their invaluable comments and suggestions, which have enhanced and bolstered the overall quality of this manuscript.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

Declarations

Conflict of interest The authors declared that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495.
- Bie, W., Fei, T., Liu, X., Liu, H., & Wu, G. (2020). Small water bodies mapped from Sentinel-2 MSI (MultiSpectral Imager) imagery with higher accuracy. *International Journal of Remote Sensing*, 41(20), 7912–7930.
- Chen, F., Chen, X., Van de Voorde, T., Roberts, D., Jiang, H., & Xu, W. (2020). Open water detection in urban environments using high spatial resolution remote sensing imagery. *Remote Sensing of Environment*, 242, 111706.
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848.
- Chen, Y., Fan, R., Yang, X., Wang, J., & Latif, A. (2018). Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water*, 10(5), 585.
- Cheng, G., Li, Z., Yao, X., Guo, L., & Wei, Z. (2017). Remote sensing image scene classification using bag of convolutional features. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1735–1739.
- Das, A., Das, S. S., Chowdhury, N. R., Joardar, M., Ghosh, B., & Roychowdhury, T. (2020). Quality and health risk evaluation for groundwater in Nadia district, West Bengal: An approach on its suitability for drinking and domestic purpose. *Groundwater for Sustainable Development*, 10, 100351.
- Dong, S., Pang, L., Zhuang, Y., Liu, W., Yang, Z., and Long, T. (2019). Optical remote sensing water-land segmentation representation based on proposed SNS-CNN network. In *IGARSS 2019–2019 IEEE international geoscience and remote sensing symposium* pp. 3895–3898.
- Du, Z., Linghu, B., Ling, F., Li, W., Tian, W., Wang, H., & Zhang, X. (2012). Estimating surface water area changes using time-series Landsat data in the Qingjiang River Basin China. *Journal of Applied Remote Sensing*, 6(1), 063609–063609.
- El-Rawy, M., Abdalla, F., & El Alfy, M. (2020). Water resources in Egypt. In *The Geology of Egypt*.
- Elsahabi, M., Negm, A., & El Tahan, A. H. M. (2016). Performances evaluation of surface water areas extraction techniques using Landsat ETM+ data: Case study Aswan High Dam Lake (AHDL). *Procedia Technology*, 22, 1205–1212.
- Enan, M. E. (2021). Deep learning for studying urban water bodies spatio-temporal transformation: a study of Chittagong City, Bangladesh (Doctoral dissertation).
- Fang, W., Wang, C., Chen, X., Wan, W., Li, H., Zhu, S., & Hong, Y. (2019). Recognizing global reservoirs from Landsat 8 images: A deep learning approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9), 3168–3177.
- Frazier, P. S., & Page, K. J. (2000). Water body detection and delineation with Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 66(12), 1461–1468.
- Gharbia, R., Hassanien, A. E., El-Baz, A. H., Elhoseny, M., & Gunasekaran, M. (2018). Multi-spectral and panchromatic image fusion approach using stationary wavelet transform and swarm flower pollination optimization for remote sensing applications. *Future Generation Computer Systems*, 88, 501–511.
- Gharbia, R., Khalifa, N. E. M., & Hassanien, A. E. (2021). Land cover classification using deep convolutional neural networks. In *Intelligent systems design and applications: 20th international conference on intelligent systems design and applications (ISDA 2020)*. Springer International Publishing.
- Guo, Q., Pu, R., Li, J., & Cheng, J. (2017). A weighted normalized difference water index for water extraction using Landsat imagery. *International Journal of Remote Sensing*, 38(19), 5430–5445.
- Gupta, A., Maheshwari, R., Guru, N., Rao, B. S., Raju, P. V., & Rao, V. V. (2022). Updated Glacial Lake inventory of Indus River Basin based on high-resolution indian remote sensing satellite data. *Journal of the Indian Society of Remote Sensing*, 12, 1–26.
- Huang, C., Chen, Y., Wu, J., Li, L., & Liu, R. (2015). An evaluation of Suomi NPP-VIIRS data for surface water detection. *Remote Sensing Letters*, 6(2), 155–164.
- Isikdogan, F., Bovik, A. C., & Passalacqua, P. (2017). Surface water mapping by deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11), 4909–4918.
- Katz, D. (2016). Undermining demand management with supply management: Moral hazard in Israeli water policies. *Water*, 8(4), 159.
- Kumari, N., Srivastava, A., & Kumar, S. (2022). Hydrological analysis using observed and satellite-based estimates: Case study of a lake catchment in Raipur, India. *Journal of the Indian Society of Remote Sensing*, 50(1), 115–128.
- Lacaux, J. P., Tourre, Y. M., Vignolles, C., Ndione, J. A., & Lafaye, M. (2007). Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal. *Remote sensing of environment*, 106(1), 66–74.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Lee, C., Kim, H. J., & Oh, K. W. (2016, October). Comparison of faster R-CNN models for object detection. In *2016 16th international conference on control, automation and systems (ICCAS)*, pp. 107–110. IEEE.
- Lin, H., Shi, Z., & Zou, Z. (2017). Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 14(10), 1665–1669.
- Mao, L., Gao, X., Zhang, Y., & Chen, Q. (2019). A comparative study of water body extraction from remote sensing images using machine learning algorithms. *International Journal of Digital Earth*, 12(7), 766–785. <https://doi.org/10.1080/17538947.2018.1520105>
- Miao, Z., Fu, K., Sun, H., Sun, X., & Yan, M. (2018). Automatic water-body segmentation from high-resolution satellite images via deep networks. *IEEE Geoscience and Remote Sensing Letters*, 15(4), 602–606.
- Noyola-Medrano, C., & Martínez-Sías, V. A. (2017). Assessing the progress of desertification of the southern edge of Chihuahuan

- Desert: A case study of San Luis Potosi Plateau. *Journal of Geographical Sciences*, 27(4), 420–438.
- Qin, X., Yang, J., Li, P., & Sun, W. (2019, July). Research on water body extraction from Gaofen-3 imagery based on polarimetric decomposition and machine learning. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 6903–6906. IEEE.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 28.
- Rokni, K., Ahmad, A., Selamat, A., & Hazini, S. (2014). Water feature extraction and change detection using multitemporal Landsat imagery. *Remote sensing*, 6(5), 4173–4189.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015*, Proceedings, Part III 18 (pp. 234–241). Springer International Publishing.
- Santoro, M., Wegmüller, U., Lamarche, C., Bontemps, S., Defourny, P., & Arino, O. (2015). Strengths and weaknesses of multi-year Envisat ASAR backscatter measurements to map permanent open water bodies at global scale. *Remote Sensing of Environment*, 171, 185–201.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- Singh, N., Nishad, R., & Singh, M. P. (2021). Automated extraction of water bodies from Landsat-8 imagery using convolutional neural network and conditional random field model. *Arabian Journal of Geosciences*, 14(3), 167.
- Tao, A., Barker, J., & Sarathy, S. (2016). Detectnet: Deep neural network for object detection in digits. *Parallel Forall*, 4.
- Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W. (2013). Selective search for object recognition. *International Journal of Computer Vision*, 104, 154–171.
- Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 41(18), 6396–6402.
- Viala, E. (2008). *Water for food, water for life a comprehensive assessment of water management in agriculture*: David Molden et al., EarthScan London and International Water Management Institute, 2007 Colombo ISBN-13: 978-1844073962.
- Wang, G., Wu, M., Wei, X., & Song, H. (2020). Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks. *Remote sensing*, 12(5), 795.
- Wang, J., Chen, K., Yang, S., Loy, C. C., & Lin, D. (2019). Region proposal by guided anchoring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (pp. 2965–2974).
- Wang, Z., Gao, X., & Zhang, Y. (2021). HA-Net: A lake water body extraction network based on hybrid-scale attention and transfer learning. *Remote Sensing*, 13(20), 4121.
- Xu, H. (2006). Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International journal of remote sensing*, 27(14), 3025–3033.
- Yan, K., Li, J., Zhao, H., Wang, C., Hong, D., Du, Y., & Wang, S. (2022). Deep learning-based automatic extraction of cyanobacterial blooms from Sentinel-2 MSI satellite data. *Remote Sensing*, 14(19), 4763.
- Yan, Y., Zhao, H., Chen, C., Zou, L., Liu, X., Chai, C., & Chen, S. (2018). Comparison of multiple bioactive constituents in different parts of *Eucommia ulmoides* based on UFLC-QTRAP-MS/MS combined with PCA. *Molecules*, 23(3), 643.
- Yang, X., & Chen, L. (2017). Evaluation of automated urban surface water extraction from Sentinel-2A imagery using different water indices. *Journal of Applied Remote Sensing*, 11(2), 026016–026016.
- Yu, Y., Huang, L., Lu, W., Guan, H., Ma, L., Jin, S., & Li, J. (2022). WaterHRNet: A multibranch hierarchical attentive network for water body extraction with remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, 115, 103103.
- Yu, J., Sharpe, S. M., Schumann, A. W., & Boyd, N. S. (2019). Detection of broadleaf weeds growing in turfgrass with convolutional neural networks. *Pest Management Science*, 75(8), 2211–2218.
- Zhang, Z., Lu, M., Ji, S., Yu, H., & Nie, C. (2021). Rich CNN features for water-body segmentation from very high-resolution aerial and satellite imagery. *Remote Sensing*, 13(10), 1912.
- Zhou, Y., Dong, J., Xiao, X., Xiao, T., Yang, Z., Zhao, G., & Qin, Y. (2017). Open surface water mapping algorithms: A comparison of water-related spectral indices and sensors. *Water*, 9(4), 256.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.