Turkish Journal of Remote Sensing, 2025; 7(1); 91-106 Research Article

Deep learning-based semantic segmentation for surface water extraction from Sentinel-2 imagery: Case study of Kuş and Uluabat lakes, Türkiye

Sohaib K. M. Abujayyab[*1]

[1] *Karabuk University, Faculty of Letters, Department of Geography, Karabuk, Türkiye;* sjayyab@karabuk.edu.tr

Article History:
Received: 26 February 2025
Revised: 14 April 2025
Accepted: 11 May 2025
Published: 30 June 2025

Abstract: This study presents a deep learning-based approach for high-precision surface water extraction from Sentinel-2 multispectral imagery. A modified U-Net architecture, trained and evaluated on two Turkish lake systems (Kuş and Uluabat Lakes), achieved superior performance compared to traditional methods. The model attained an overall accuracy of 0.9980, precision of 0.9980, recall of 0.9980, F1-score of 0.9980, and Intersection over Union (IoU) of 0.9961, outperforming both Normalized Difference Water Index (NDWI) and Modified NDWI (MNDWI). Analysis reveals that the U-Net effectively mitigates spectral confusion in heterogeneous environments, demonstrating its potential for enhanced water resource monitoring, flood mapping, and hydrological modeling applications. While NDWI and MNDWI achieved IoU scores of 0.9956 and 0.9953, respectively, the deep learning model's higher IoU signifies more accurate boundary delineation. The improved performance highlights the value of deep learning in automated surface water mapping for enhanced decision-making in water resource management. These results suggest that while traditional spectral indices are useful for preliminary analysis, deep learning approaches offer a more refined classification, particularly in complex or heterogeneous landscapes.

Keywords: Surface water extraction; deep learning; semantic segmentation; Sentinel-2; Kuş and Uluabat lakes

Citation: Abujayyab, S. K. M. (2025). Deep learning-based semantic segmentation for surface water extraction from Sentinel-2 imagery: Case study of Kuş and Uluabat lakes, Türkiye. *Turk. J. Remote Sens.,* 7(1), 91-106. https://doi.org/10.51489/tuzal.1647078

1. Introduction

Surface Water, encompassing lakes, rivers, reservoirs, and coastal areas, are integral components of the Earth's complex ecosystem. Their influence extends far beyond their physical boundaries, impacting regional and global climate patterns, sustaining diverse aquatic and terrestrial habitats, and providing essential resources for human populations (Al-Najjar et al., 2019). Consequently, accurate, reliable, and up-to-date information on the spatial extent, distribution, and dynamics of water resources is paramount for informed decision-making in a multitude of applications, including water resource management, flood risk assessment, ecological monitoring, and land use planning (Yilmaz, 2023).

Remote sensing technologies, such as Sentinel-2 multispectral imaging, provide high-resolution spatial and temporal data, enabling large-scale analysis of water dynamics. However, traditional methods for water body extraction often rely on manual thresholding or spectral indices such as the Normalized Difference Water Index (NDWI) (Moumane et al., 2025)and the Modified Normalized Difference Water Index (MNDWI) (Yao et al., 2025), which exploit the strong absorption of water in the near-infrared (NIR) and shortwave infrared (SWIR) regions of the electromagnetic spectrum. While these spectral indices offer a simple and computationally efficient approach for water body mapping, they can be susceptible to errors caused by spectral confusion with other land cover types, such as shadows, vegetation, and built-up areas.

Supervised classification techniques, such as maximum likelihood and support vector machines (SVMs) (Liong & Sivapragasam, 2002; Pang et al., 2023), have also been widely used for water body extraction from remote sensing imagery. These methods require training data, which consists of labeled samples representing different land cover classes, and can achieve high accuracy when trained with representative data (Li et al., 2021). However, the performance of supervised classification methods is highly dependent on the quality and quantity of training data, and they can struggle to generalize well to unseen areas or time periods. Furthermore, these traditional methods often require manual feature engineering, where domain experts select and extract relevant spectral, spatial, and textural features from the imagery to improve classification accuracy. This manual feature engineering process can be time-consuming, subjective, and may not always capture the full complexity of the underlying data.

In recent years, the field of image analysis has been transformed by the rapid development and widespread adoption of deep learning (DL) techniques (Digra et al., 2022; Shabbir et al., 2021). DL, a subfield of machine learning, employs artificial neural networks with multiple layers to learn hierarchical representations of data, enabling the automatic extraction of complex features from raw input data. Deep learning architectures, particularly convolutional neural networks (CNNs), have achieved remarkable success in a wide range of computer vision tasks, including image classification, object detection, and semantic segmentation. Semantic segmentation, in particular, has emerged as a powerful tool for pixel-wise classification of images, assigning a specific category label to each pixel in an image. This fine-grained classification capability makes semantic segmentation ideally suited for extracting complex spatial features from remote sensing imagery, such as water bodies, with high accuracy and detail (Al-Najjar et al., 2019; Digra et al., 2022; Zhao et al., 2020).

The application of DL architectures, such as convolutional neural networks (CNNs), has demonstrated remarkable success in remote sensing analysis. For instance, studies leveraging CNNs for LULC classification have achieved superior accuracy in distinguishing fine-grained features, including buildings, vegetation, and water bodies, by integrating multi-modal data like digital surface models (DSMs) (Al-Najjar et al., 2019). Similarly, residual dense networks (RDNs) enhanced with channel-spatial attention mechanisms have shown improved feature representation in high-resolution imagery (Zhao et al., 2023). Transfer learning approaches, such as fine-tuning pre-trained models like ResNet50 and Inception-v3, further optimize performance while reducing computational costs (Fayaz et al., 2024; Shabbir et al., 2021). Despite these advancements, surface water extraction poses unique challenges due to the spectral similarity between water and shadows, seasonal variations, and the need for precise boundary delineation in Sentinel-2 data.

While existing frameworks like DPPNet Priyanka et al. (2023) and U-Net (Atik, 2023) have advanced semantic segmentation in remote sensing, few studies explicitly address water body extraction using Sentinel-2's spectral bands. Campos-Taberner (2023) emphasized the interpretability of DL models for Sentinel-2 time series, identifying red and near-infrared bands as critical for LULC tasks (Campos-Taberner et al., 2020). However, the integration of multi-spectral data with deep semantic segmentation remains underexplored for water-specific applications.

This research focuses on the application of deep learning-based semantic segmentation for extracting surface water from Sentinel-2 imagery. By leveraging the power of CNNs, this research aims to develop a robust and efficient method for automatically mapping surface water with high accuracy and detail from Sentinel-2 imagery.

2. Materials and Methods

2.1. Study area

This study focuses on two significant lake systems in Turkey: Kuş and Uluabat Lakes, both located within the Marmara region and exhibiting contrasting ecological characteristics (See Figure 1.)(Barlas et al., 2005; Filik Iscen et al., 2008; Uzun, 2024). Kuş Lake, a relatively small, shallow freshwater lake situated near Bandırma, is recognized as an important wetland ecosystem and a Ramsar site, supporting a diverse array of bird species, particularly during migration seasons. Its ecological importance stems from its role as a key stopover point for migratory birds traveling along the western Palearctic flyway. The lake's shallow depth and surrounding reed beds provide essential habitat for nesting, foraging, and resting. However, Kuş Lake faces increasing pressures from agricultural runoff, industrial discharge, and climate change, impacting water quality and threatening the delicate balance of its ecosystem (Dervisoglu, 2021; Yilmaz, 2023).

In contrast, Uluabat Lake, is a much larger and deeper freshwater lake located south of Bursa. As designated Ramsar site, it holds significant ecological value, supporting a rich biodiversity of fish, birds, and other aquatic life. Uluabat Lake plays a critical role in regional water regulation and supports local fisheries. However, similar to Kuş Lake, it faces growing environmental challenges, including pollution from surrounding agricultural activities, industrial development, and urban expansion. These pressures contribute to eutrophication, habitat degradation, and the decline of native species.

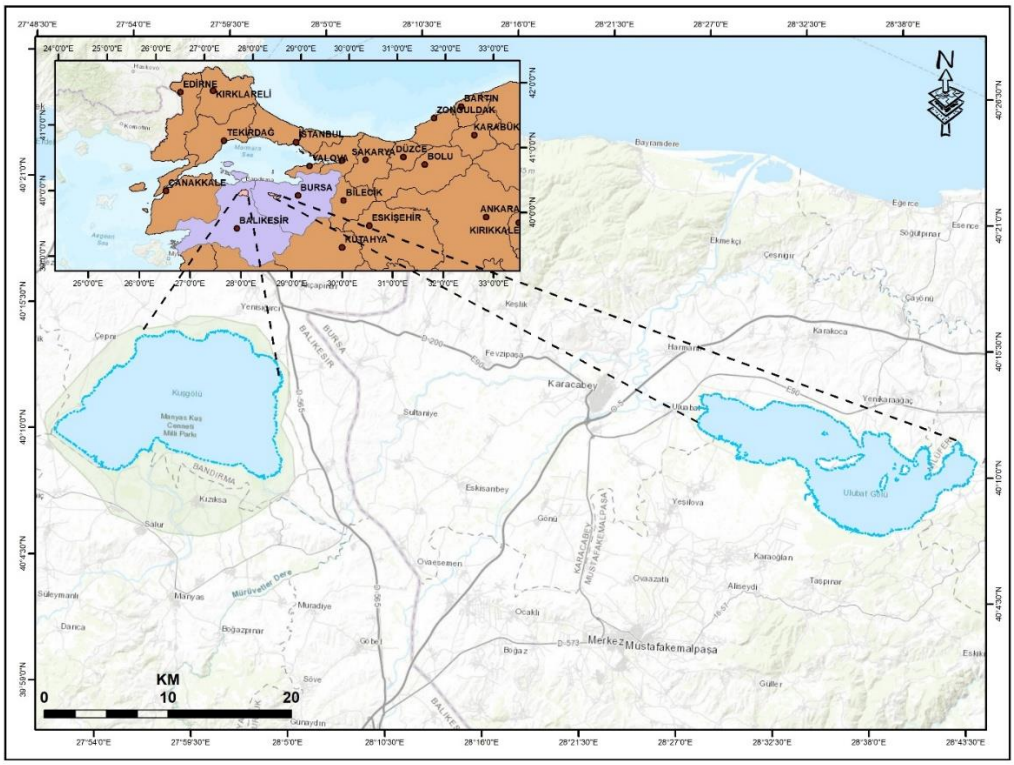

Figure. 1. Location of the study area Kuş (Training) and Uluabat Lakes (Testing) in Turkey.

The selection of these two lakes provides a valuable comparative context for evaluating the performance of the proposed surface water extraction method. Their contrasting characteristics, in terms of size, depth, surrounding land use, and ecological pressures, present diverse scenarios for assessing the model's ability to accurately delineate water boundaries and monitor changes in water extent.

2.2. Data collection

Sentinel-2 Surface Reflectance imagery, provided by the European Space Agency (ESA) via the Copernicus program, was utilized in the study area. The Uluabat and Kuş lakes were considered as areas of interest (AOIs). The AOIs were delineated using latitude-longitude coordinates in the Google Earth Engine (GEE) platform.

The dataset was filtered for the period between January 1, 2020, and December 30, 2020, to ensure temporal consistency. To minimize the impact of atmospheric disturbances, images with less than 20% cloud cover were pre-selected.

The selected bands included the visible spectrum (B4, B3, B2) for red, green, and blue, respectively, and the near-infrared band (B8) to enhance vegetation analysis. Median composite imagery was generated by aggregating all cloud-free images within the specified period, ensuring a representative dataset for subsequent analysis.

2.3. Dataset preprocessing

The preprocessing of the dataset was conducted to ensure the quality, consistency, and readiness of the data for deep learning-based water bodies mapping. The raw satellite imagery and corresponding ground truth labels underwent a series of transformations, including handling invalid values, normalization, and tile generation, to prepare the data for model training and testing.

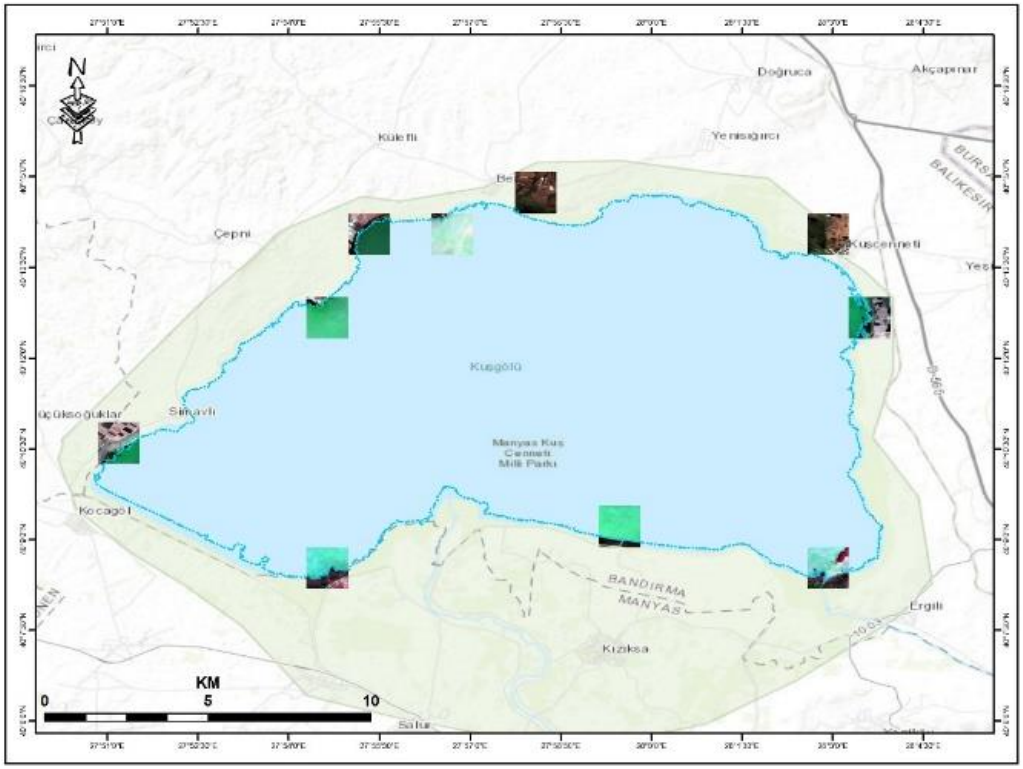

Figure. 2. Locations of an example of the training Sentinel-2 images

Initially, Sentinel-2 imagery and binary water Label were clipped to the region of interest (ROI) and tiled into smaller sections of 128×128 pixels using the GeoTile library (See Figure 2.). This tiling approach ensured the compatibility of the data with the input size requirements of the deep learning model and optimized memory usage. Each tile was carefully labeled with its respective water label to establish a supervised learning dataset.

Next, normalization was applied to the pixel values of the Sentinel-2 imagery, scaling them to a range of 0 to 1. This step was crucial for enhancing model convergence by standardizing input values and eliminating the influence of varying data scales.

A validation process was carried out to identify and exclude erroneous tiles. Specifically, label tiles containing invalid values, such as pixel intensities exceeding the binary range (0 and 1), were flagged, and removed. This step ensured that only valid and meaningful tiles contributed to the model training process. After validation, the dataset consisted of 600 high-quality tiles, each containing multispectral Sentinel-2 imagery and its corresponding binary water label. Figure 3. Illustrating few examples from the RGB Sentinel-2 imagery and thir corresponding label for each image.

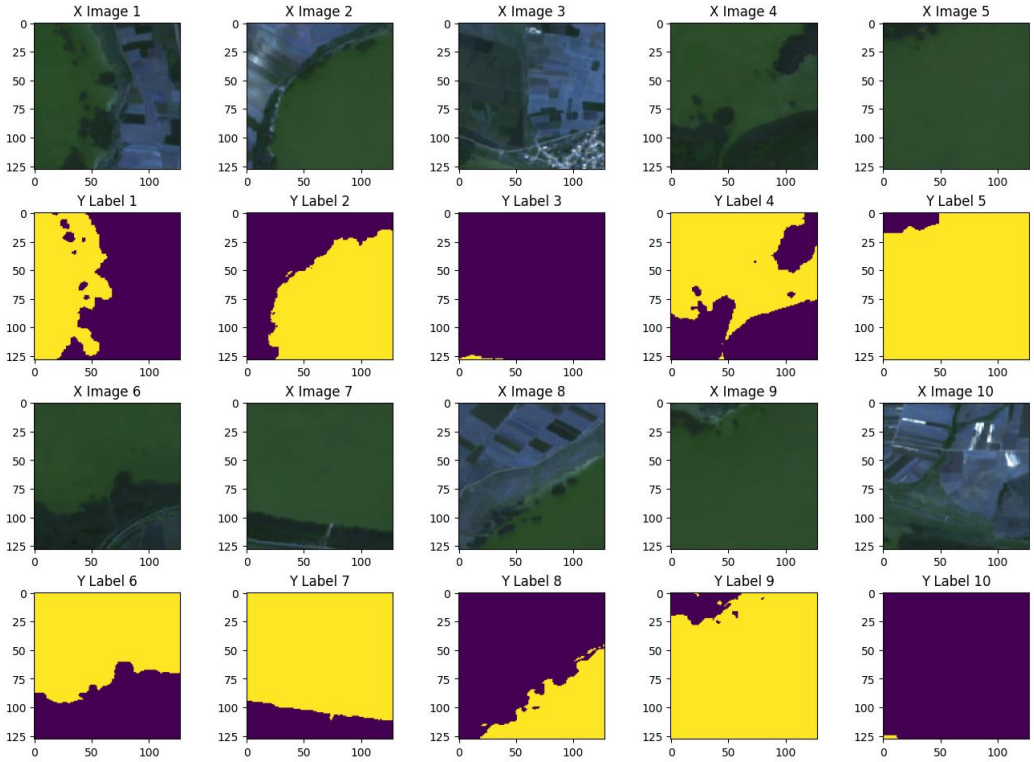

Figure. 3. An example of the training Sentinel-2 tiles including (B4, B3, B2 bands) for red, green, and blue, respectively, and the near-infrared band (B8) with their respected water Label

2.4. U-Net model development

In this study, a U-Net architecture was developed for surface water mapping (See Figure 4.), leveraging its strengths in semantic segmentation tasks. The model was designed to process input tiles of size 128×128×4, where the four input channels correspond to the multispectral bands of Sentinel-1 or Sentinel-2 imagery. The U-Net model consists of two main components: an encoder for feature extraction and a decoder for reconstructing spatial information, with skip connections linking the two to retain fine details.

The encoder begins with an input layer that accepts tiles of shape (128, 128, 4). This input passes through a series of convolutional layers that apply a 3×3 kernel with ReLU activation and same padding to preserve spatial dimensions. Dropout layers are employed after each convolution to prevent overfitting by randomly deactivating a fraction of neurons. The first convolutional block extracts 32 feature maps, followed by a max-pooling operation to reduce spatial dimensions by half. This process is repeated, with the number of feature maps increasing to 64 in deeper layers, enhancing the model's capacity to capture complex spatial patterns.

The final encoding layer produces a compact representation of the input, balancing spatial resolution and feature richness.

The decoder reconstructs the spatial dimensions of the input through transposed convolutional layers, which perform upsampling while learning reconstruction filters. Skip connections from the encoder are concatenated with the corresponding decoder layers to recover spatial details lost during downsampling. Each decoder block comprises transposed convolutions with ReLU activation and dropout layers to maintain regularization. The decoder progressively increases the spatial resolution back to the original tile size while reducing the number of feature maps.

To generate the binary for water body map, the final decoder output passes through a series of 1×1 convolutional layers, which refine the output to a single channel. The final activation function is a sigmoid, suitable for binary classification tasks, as it maps pixel intensities to values between 0 and 1.

The U-Net model was compiled using the Adam optimizer, which provides adaptive learning rates for efficient training. The binary cross-entropy loss function was used, given the binary nature of the output (Water or no-Water), along with accuracy as an evaluation metric. This U-Net implementation effectively balances spatial detail retention and computational efficiency, making it well-suited for water bodies mapping.

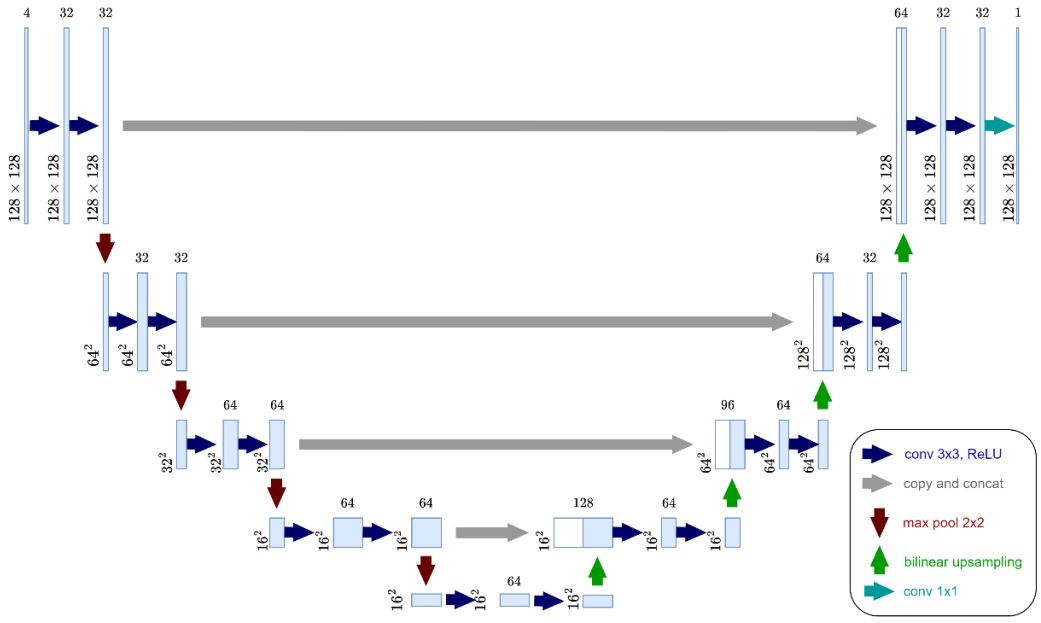

Figure. 4. The architecture of the U-Net.

To train the U-Net model for water bodies mapping, TensorFlow's deep learning framework were utilized. The training process involved the use of callbacks and hyperparameter tuning to ensure optimal model performance and prevent overfitting. A key component of the training strategy was the integration of a checkpointing mechanism using TensorFlow's ModelCheckpoint callback. This callback monitored the validation loss during training and saved the model weights only when the validation loss improved, ensuring that the best-performing model was preserved. The callback function was configured with the "val_loss" metric as the monitor, a verbose output to track progress, and a "min" mode to identify and save the model with the minimum validation loss. The use of ModelCheckpoint mitigated the risk of overfitting, as it allowed the training process to store optimal weights while discarding suboptimal training iterations. This ensured that the final model was generalizable to unseen data. An optional early stopping mechanism, which halts training if

the validation loss does not improve for a specified number of epochs, was considered but not activated for this particular implementation to allow the model to train for the full 50 epochs. The model was trained on the preprocessed dataset using 50 epochs and a batch size of 16.

The model utilizes the Adam optimizer with its default configuration during training. Specifically, the learning rate is set to 0.001, which is the standard default value for Adam in TensorFlow. This adaptive optimizer is well-suited for binary segmentation tasks due to its efficient handling of sparse gradients and noisy data.

During training, the model's performance was monitored in terms of the loss values for both training and validation datasets. By saving the model weights at the point of lowest validation loss, the risk of overfitting or underfitting was minimized, and the saved model could then be used for subsequent evaluation and prediction tasks.

This training framework ensured the development of a robust U-Net model optimized for accurate surface water mapping, capable of effectively generalizing to new spatial data.

2.5. Performance evaluation

The performance of the proposed deep learning model for surface water extraction was assessed using a suite of established metrics: Precision, Recall, F1-score, and Intersection over Union (IoU). These metrics provide a comprehensive evaluation of the model's ability to correctly classify water pixels while minimizing both false positives (non-water pixels classified as water) and false negatives (water pixels misclassified as non-water). This multi-faceted evaluation allows for a nuanced understanding of the model's strengths and weaknesses in accurately delineating surface water from Sentinel-2 imagery.

Precision, quantifying the accuracy of positive predictions, measures the proportion of correctly identified water pixels out of all pixels classified as water by the model. A high precision indicates a low rate of false positives, meaning the model is effectively avoiding misclassifying non-water features as water.

$$Precision = \frac{TP}{(TP + FP)} \quad (1)$$

Recall, also known as sensitivity, assesses the model's ability to capture all actual water pixels. It measures the proportion of correctly identified water pixels out of all true water pixels present in the ground truth data. A high recall signifies a low rate of false negatives, indicating the model is successfully identifying the vast majority of water pixels within the study area.

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

The F1-score, the harmonic mean of precision and recall, provides a balanced measure of overall accuracy, particularly valuable when dealing with potentially imbalanced datasets where one class (e.g., water) may have significantly fewer pixels than another.

$$F1 - score = 2\,x\,\frac{(Precision\ x\ Recall)}{(Precision + Recall)} \quad (3)$$

Finally, the Intersection over Union (IoU), offers a direct measure of the spatial overlap between the predicted water mask and the ground truth water mask. It is calculated as the ratio of the intersection of the predicted and ground truth masks to the union of the two masks. A higher IoU signifies better segmentation performance, indicating a closer alignment between the model's predicted water boundaries and the actual water extent.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

A represents the set of pixels in the predicted segmentation mask. B represents the set of pixels in the ground truth mask.

|A ∩ B| denotes the number of pixels in the intersection of A and B (i.e., the number of true positives). It represents the area where both the prediction and the ground truth agree that the pixel belongs to the object of interest (in this case, water).

|A ∪ B| denotes the number of pixels in the union of A and B (i.e., the sum of true positives, false positives, and false negatives). It represents the total area covered by either the prediction or the ground truth, or both.

3. Results

This section presents the results of the deep learning-based approach for water body extraction using Sentinel-2 imagery. The performance of the U-Net model evaluated using several metrics, including Precision, Recall, F1-score, and IoU. a visual comparison of predicted water bodies against actual boundaries provided. The results are further discussed in comparison with traditional models.

3.1. Analysis of the training log

The U-Net model's performance was assessed using a suite of metrics designed to capture its effectiveness in water body delineation. The model was trained for 20 epochs. As shown in Figure 5, both the training and validation accuracy increased over the epochs, while the corresponding loss values decreased. This trend indicates consistent improvement in the model's ability to distinguish surface water from other land cover types. The use of the ModelCheckpoint callback used to save the best model weights.

The model achieved a final training accuracy of 0.9990 and a validation accuracy of 0.9976. Due to the high quality of the satellite images and low noise, the slight difference between training and validation accuracy indicates a minor degree of overfitting. However, the high validation accuracy still demonstrates the model's robustness and its capacity to perform well on unseen data. Examining the loss curves in Figure 5 reveals that the training loss stabilized around 0.0026, while the validation loss reached its lowest point of 0.0078 at epoch 20. Each epoch took between 170 and 201 seconds. Therefore, the total training duration is 3487 seconds. The early stabilization of the validation loss suggests that further training beyond this point might offer limited performance gains and could potentially increase the risk of overfitting.

The rapid convergence of the model, achieving near-optimal performance within a few epochs, as observed in Figure 5, can indeed be unusual in some deep learning scenarios. However, several factors contribute to this rapid convergence: (A) The Sentinel-2 imagery used in this study is of high quality with minimal noise. The preprocessing steps, including cloud masking and atmospheric correction, further enhance the data quality and reduce potential confounding factors. This clean and consistent input data facilitates efficient model training and faster convergence. (B)The U-Net architecture, with its encoder-decoder structure and skip connections, is particularly well-suited for semantic segmentation tasks. Its design enables efficient capture of both local and global contextual information, leading to faster learning and convergence, especially for well-defined features like water bodies. (C) The inclusion of the near-infrared (NIR) band, alongside the red, green, and blue (RGB) bands, significantly contributes to the rapid convergence. Water exhibits strong absorption in the NIR spectrum, providing a distinct spectral signature that contrasts sharply with most other land cover types. This clear spectral separability allows the model to quickly learn the differentiating characteristics of water bodies, leading to faster convergence.

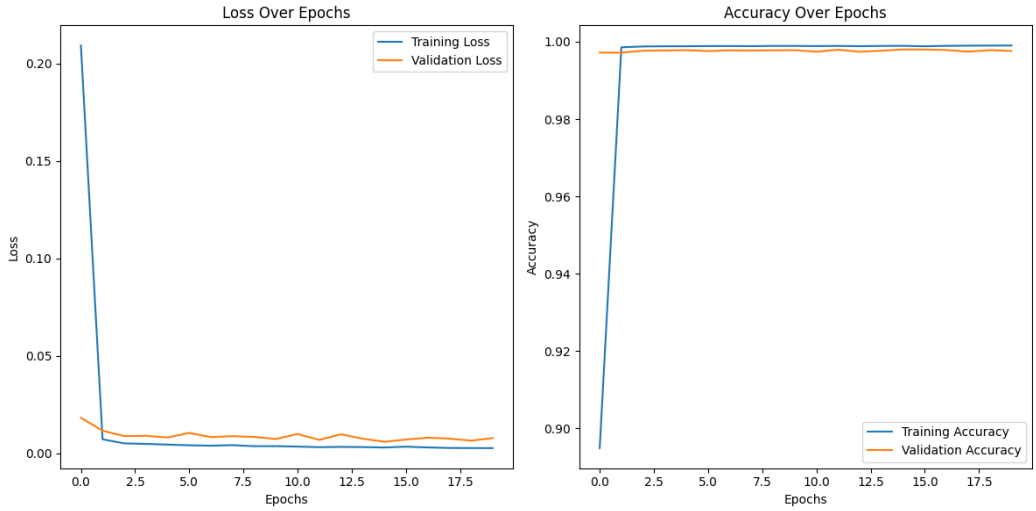

Figure.5. Training and validation accuracy (left figure), training and validation loss (right figure)

The confusion matrix and ROC curve, presented in Figure 6, offers a better view of the model's classification performance. It shows 1474835 true positives (correctly identified water pixels), 3198 true negatives (correctly identified non-water pixels), 8336146 false positives (non-water pixels misclassified as water), and 16221 false negatives (water pixels misclassified as non-water). Based on these values, the calculated Precision is 0.9992, Recall is 0.9851, F1-score is 0.9921, and IoU is 0.9843. The high precision indicates a low rate of false positives, meaning the model effectively avoids misclassifying non-water features as water. The slightly lower recall, while still high, suggests a small number of missed water pixels. The F1-score, a balanced measure of precision and recall, confirms the model's overall strong performance, and the IoU further corroborates this by quantifying the significant overlap between predicted and actual water body extents. The ROC curve with an AUC of 0.99 highlights the model's excellent discrimination capability, with a low false positive rate. These results suggest that the model successfully generalizes to unseen data, which is crucial for large-scale water body mapping.

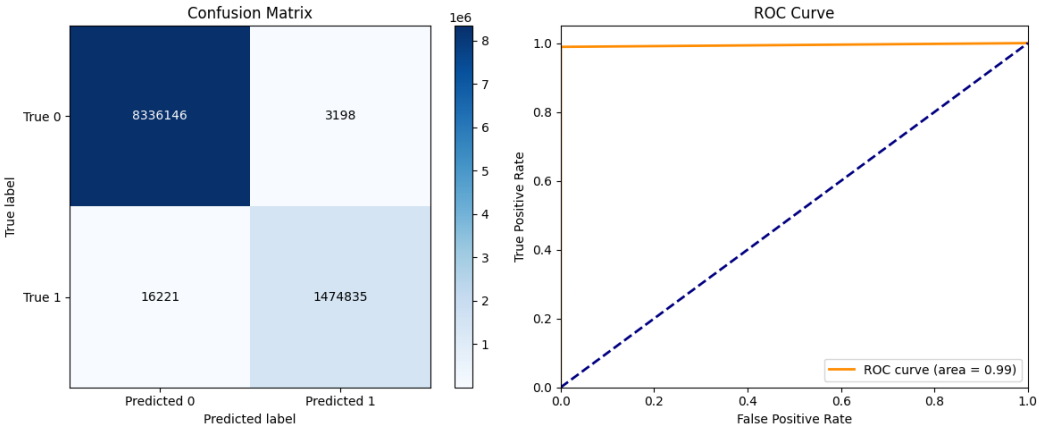

Figure.6. Model performance evaluation: confusion matrix and ROC curve

Analysis of the training log reveals fluctuations in the time taken per epoch. This variation is likely due to external factors such as system load and does not reflect inherent model inefficiency. While the training time varied from 171 seconds to as high as 201 seconds, the average time per epoch gives a better representation of the model's computational needs. This information can be valuable in estimating resource requirements

for future applications or for comparing the efficiency of different model architectures. The detailed breakdown of performance metrics provides a comprehensive understanding of the model's strengths and limitations, highlighting its effectiveness in accurately mapping surface water from Sentinel-2 imagery.

3.2. Comparison of predicted water bodies against actual boundaries

The comparison between the predicted water bodies and the actual ground truth demonstrates the high accuracy of the model in delineating surface water. The input image, ground truth, and predicted output are visualized side by side in Figure.7. The predicted water extent closely matches the ground truth, with minimal contradictions along the boundary regions.

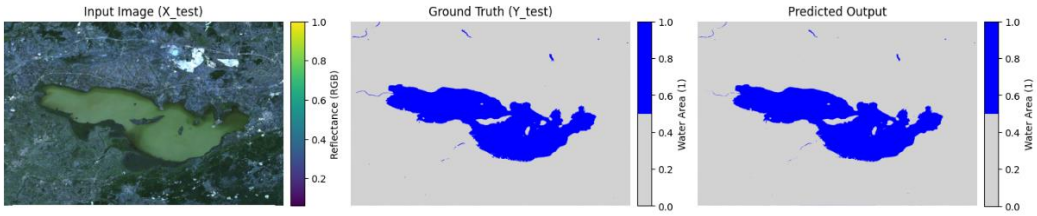

Figure.7. Water body segmentation: model prediction vs. ground truth

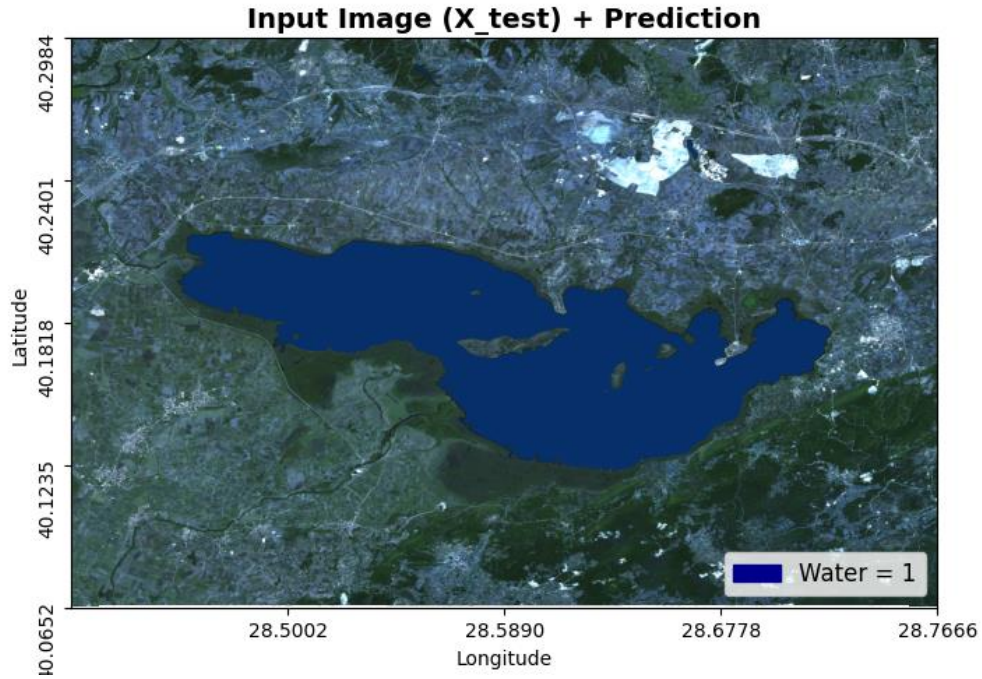

Figure.8. Predicted water area overlaid on original Sentinel 2 Satellite Image.

The predicted water body is overlaid on the input image (See Figure 8.), offering a spatial perspective of the classification. The predicted water extent aligns well with the lake's actual boundaries, demonstrating the model's ability to accurately segment surface water in satellite imagery. However, minor deviations may be observed along the edges, potentially due to spectral similarities between water and adjacent land cover types, such as wetlands or shadowed areas. This analysis highlights the effectiveness of the model in surface water detection while also emphasizing the need for further refinement in complex boundary regions. The findings support the potential application of this approach in hydrological studies, environmental monitoring, and flood mapping, where precise water body delineation is essential.

3.3. Comparison with traditional models

To assess the effectiveness of the proposed deep learning approach, the performance of traditional surface water extraction methods compared. Specifically, U-Net model benchmarked against two widely used spectral indices: the NDWI and MNDWI. These indices exploit the spectral characteristics of water in the near-infrared (NIR) and shortwave infrared (SWIR) regions of the electromagnetic spectrum to distinguish water from other land cover types.

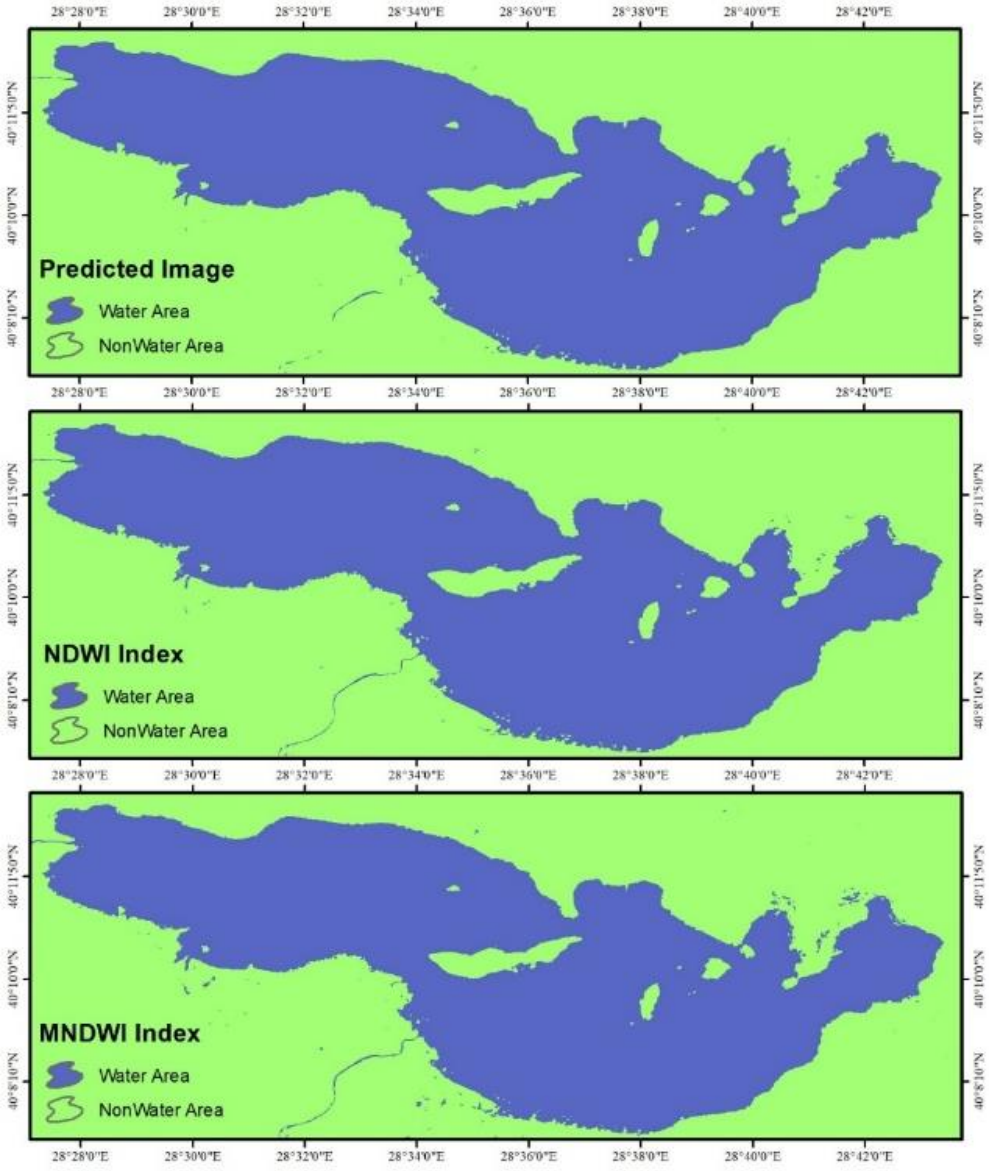

Figure.9. Comparison of predicted water areas with NDWI, and MNDWI

The overall agreement between the predicted image and the spectral indices highlights the robustness of the model in surface water delineation. However, minor discrepancies along complex shoreline regions suggest that while the model effectively generalizes water features, it may still require refinement in areas with mixed land-water characteristics. The advantage of the deep learning-based prediction is its ability to integrate spatial and contextual information beyond simple spectral thresholding, making it more adaptable in challenging environments such as turbid or shallow waters. The results indicate that while traditional spectral indices remain useful for water mapping, machine learning-based

approaches offer promising improvements in classification accuracy, especially in heterogeneous landscapes where spectral confusion is prevalent.

The confusion matrices (See Figure 10.) provide a comparative evaluation of the performance of the deep learning-based prediction model against the NDWI and MNDWI. The predicted image achieves the lowest false positives (4,016) and false negatives (16,277), demonstrating a strong ability to distinguish between water and non-water areas. In contrast, the NDWI method exhibits a slightly higher number of false negatives (18,975), suggesting it may under-classify some water pixels, potentially due to variations in water turbidity or spectral similarities with land surfaces. The MNDWI index, while effective, records the highest false positives (16,553), indicating a tendency to misclassify certain land areas as water, which may be influenced by factors such as urban features or wet soil reflecting similarly to water in the modified spectral bands.

Table 1. Comparison of Classification Performance: Predicted Image vs. NDWI and MNDWI

	Accuracy	Precision	Recall	F1-score	IoU
Predicted Image	0.9980	0.9980	0.9980	0.9980	0.9961
NDWI	0.9978	0.9978	0.9978	0.9978	0.9956
MNDWI	0.9976	0.9976	0.9976	0.9976	0.9953

The performance metrics in the Table.1. demonstrate that the predicted image from the model achieves the highest accuracy (0.9980), precision (0.9980), recall (0.9980), F1-score (0.9980), and Intersection over Union (IoU = 0.9961) compared to the NDWI and MNDWI indices. While all three approaches exhibit strong performance, the deep learning-based prediction slightly outperforms the NDWI and MNDWI, particularly in terms of IoU, indicating better spatial agreement with the reference surface water. The NDWI and MNDWI indices, which rely on spectral thresholding techniques, show very similar performance but with marginally lower scores than the model's prediction. This suggests that while spectral indices remain effective for surface water delineation, machine learning-based methods offer superior precision and generalization, especially in complex environments where spectral misclassification may occur.

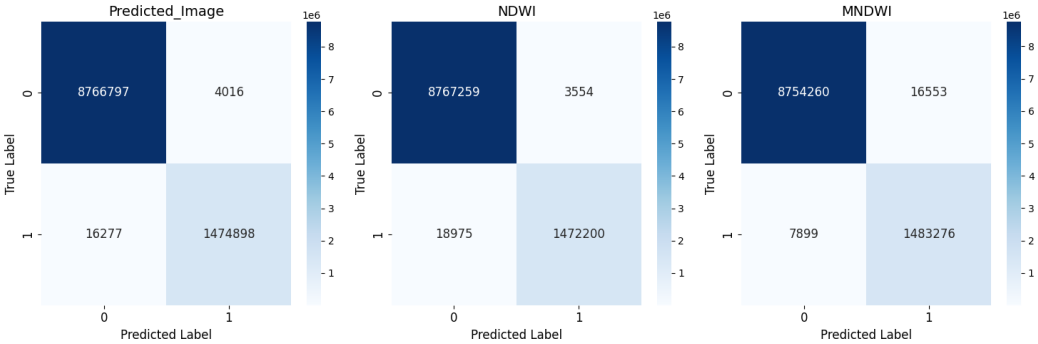

Figure.10. Comparison of confusion matrices for predicted water bodies, NDWI, and MNDWI classification

The consistently high performance across all methods highlights the reliability of satellite-based surface water mapping. However, the slight differences between the predicted image and spectral indices suggest that traditional thresholding methods might misclassify certain land-water boundary regions, particularly in areas with mixed land cover, turbid water, or shadow effects. The deep learning model, on the other hand, integrates spatial and contextual information, leading to improved segmentation results.

4. Discussion

In evaluating surface water extraction methodologies, it is vital to compare deep learning-based optical image processing, specifically using U-Net models with Sentinel-2 data, against Synthetic Aperture Radar (SAR)-based approaches. This discussion reveals insights into their distinctive advantages and operational contexts. Deep learning models like U-Net have gained prominence in optical image processing due to their superior performance in extracting fine spatial details. For instance, U-Net architectures demonstrate high accuracy when applied to complex image segments such as water bodies, benefiting from high spectral resolution data available in Sentinel-2 imagery(Solórzano et al., 2021). Their ability to learn intricate patterns through convolutional layers enables a nuanced understanding of surface features, leading to precise segmentation.

Conversely, SAR-based water extraction offers unique advantages under conditions unfavorable for optical imagery, such as cloud cover or poor lighting, where optical sensors struggle. The utility of SAR data lies in its all-weather, day-and-night imaging capability, making it invaluable for consistent surface water monitoring (Bioresita et al., 2018; Xie et al., 2023). Prior research highlights the efficacy of traditional SAR methods, such as those employing the Automatic Water Mapping (AWM) algorithm, which yield high accuracy in various applications (Bioresita et al., 2018). Moreover, studies highlight the precision of using SAR imagery for continuous water monitoring in challenging climatic regions like the Mekong Delta (Pham-Duc et al., 2017).

In direct comparisons, multiple studies underscore the efficiency of SAR processing algorithms, demonstrating comparable or superior performance against optical indices. For instance, Zhang et al. (2020) applied a robust thresholding technique specifically adapted for SAR data, achieving reliable surface water mapping that outperformed analog optical methods. Furthermore, the fusion of Sentinel-1 and Sentinel-2 imagery is explored, where the combined datasets can synergistically enhance water detection, allowing for the extraction of more complex features (Bai et al., 2021).

Literature has pointed out that while traditional SAR algorithms can be computationally efficient, they often require significant parameter tuning and understanding of local geographic features to optimize their performance—facets where deep learning models might provide improved usability and scalability due to their learning capabilities from large datasets (Šćepanović et al., 2021). This adaptability makes U-Net a compelling choice, especially in specific applications such as flood mapping and hydrological modeling, where rapid and precise data analysis is necessary (Asaro et al., 2021).

Therefore, while both methodologies have their merits, the choice between deep learning optical image processing and SAR-based techniques hinges on the specific use case scenarios, environmental conditions, and operational requirements. The deployment of both methods can also complement each other, enhancing overall accuracy and reliability in surface water extraction tasks.

5. Conclusions

This research demonstrates the power of deep learning for automated water body extraction from Sentinel-2 imagery, offering a significant advancement over traditional methods. The modified U-Net model's ability to accurately delineate water boundaries, even in complex environments, has broad implications for water resource management, environmental monitoring, and ecological studies.

The modified U-Net model consistently outperformed traditional spectral index-based methods, achieving higher precision, recall, F1-score, and IoU. This improvement stems from the model's ability to learn complex spatial and spectral patterns, effectively distinguishing water from other land cover types, even in challenging scenarios with mixed land-water

interfaces, turbidity, or shadow effects. The deep learning approach automates the water body extraction process, significantly reducing manual effort and enabling large-scale analysis of water resources. This efficiency is crucial for timely monitoring and informed decision-making. The model's consistent performance across two distinct lake systems suggests good generalizability to diverse hydrological and ecological contexts. This robustness makes the approach applicable to various geographic regions and water body types.

The accurate and automated delineation of water bodies using this deep learning approach has far-reaching Broader implications for:

- Water Resource Management: Provides critical information for water allocation, drought monitoring, and sustainable water use planning, ultimately contributing to more effective water resource management.
- Hydrological Modeling: Improves the accuracy of hydrological models by providing precise water extent data, enabling better predictions of water availability, flow dynamics, and flood risks.
- Flood Risk Assessment and Mitigation: Supports the development of accurate flood inundation maps, contributing to better risk assessments and the design of effective flood mitigation strategies.
- Environmental Monitoring and Ecological Studies: Facilitates the monitoring of changes in water bodies over time, enabling the detection of trends related to climate change, land use changes, pollution, and other environmental pressures. This information is crucial for understanding ecosystem health and supporting conservation efforts.

Researchers can build upon this work by exploring several promising avenues:

- Extend the analysis to multi-temporal Sentinel-2 data to investigate water body dynamics, including seasonal variations, long-term trends, and the impact of climate change. Develop specific change detection algorithms for water bodies and use time-series analysis to understand evolving patterns. This could involve incorporating recurrent neural networks or long short-term memory networks to handle the temporal dimension.
- Investigate the fusion of Sentinel-2 data with other satellite sensors (e.g., Landsat-9, PlanetScope) to enhance spatial resolution, improve detection of smaller or fragmented water bodies, and address spectral confusion in challenging environments. Research could focus on optimal fusion strategies and comparative analysis of their effectiveness.
- Explore advanced deep learning architectures and techniques, such as attention mechanisms or transformers, to further refine model accuracy. Develop methods to quantify the uncertainty associated with model predictions to enhance reliability assessment and support informed decision-making. Bayesian deep learning approaches could be a valuable area of exploration.
- Integrate the deep learning-based water body extraction method with hydrological models to improve model calibration, validation, and predictive capabilities. This would contribute to more accurate and reliable water resource assessments.
- A key area for future research is analyzing the model's robustness across varying hydrological conditions. This involves using a multi-year dataset encompassing distinct dry and rainy seasons to: (1) train and evaluate the model separately on data from these periods to assess performance differences and identify specific challenges, and (2) develop a more generalized model by incorporating temporal information or training on a more diverse dataset that captures the full range of variability in water levels and quality. This will lead to a more comprehensive understanding of the model's limitations and contribute to the development of more reliable deep learning-based approaches for water body extraction.

Author Contributions: A single author carried out the study.

Research and publication ethics statement: In the study, the author declares that there is no violation of research and publication ethics and that the study does not require ethics committee approval.

Conflicts of Interest: The author declares no conflict of interest.

References

Al-Najjar, H. A. H., Kalantar, B., Pradhan, B., Saeidi, V., Halin, A. A., Ueda, N., & Mansor, S. (2019). Land cover classification from fused DSM and UAV images using convolutional neural networks. *Remote Sensing*, 11(12). https://doi.org/10.3390/rs11121461

Asaro, F., Murdaca, G., & Prati, C. M. (2021, July 11-16). Learning deep models from weak labels for water surface segmentation in Sar images [Paper presentation]. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium. https://doi.org/10.1109/IGARSS47720.2021.9554647

Atik, Ş. Ö. (2023). Object-Based Integration Using Deep Learning and Multi-Resolution Segmentation in Building Extraction from Very high resolution satellite imagery. *Turkish Journal of Remote Sensing*, 5(2), 67–77. https://doi.org/10.51489/tuzal.1337656

Bai, Y., Wu, W., Yang, Z., Yu, J., Zhao, B., Liu, X., Yang, H., Mas, E., & Koshimura, S. (2021). Enhancement of detecting permanent water and temporary water in flood disasters by fusing Sentinel-1 and Sentinel-2 imagery using deep learning algorithms: Demonstration of sen1floods11 benchmark datasets. *Remote Sensing*, 13(11). https://doi.org/10.3390/rs13112220

Barlas, N., Akbulut, N., & Aydoğan, M. (2005). Assessment of heavy metal residues in the sediment and water samples of Uluabat Lake, Turkey. *Bulletin of Environmental Contamination and Toxicology*, 74(2), 286–293. https://doi.org/10.1007/s00128-004-0582-y

Bioresita, F., Puissant, A., Stumpf, A., & Malet, J.-P. (2018). A method for automatic and rapid mapping of water surfaces from Sentinel-1 imagery. *Remote Sensing*, 10(2). https://doi.org/10.3390/rs10020217

Campos-Taberner, M., García-Haro, F. J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G., & Gilabert, M. A. (2020). Understanding deep learning in land use classification based on Sentinel-2 time series. *Scientific Reports*, 10(1), 17188. https://doi.org/10.1038/s41598-020-74215-5

Dervisoglu, A. (2021). Analysis of the temporal changes of inland ramsar sites in Turkey using google earth engine. *ISPRS International Journal of Geo-Information*, 10(8). https://doi.org/10.3390/ijgi10080521

Digra, M., Dhir, R., & Sharma, N. (2022). Land use land cover classification of remote sensing images based on the deep learning approaches: a statistical analysis and review. *Arabian Journal of Geosciences*, 15(10), 1003. https://doi.org/10.1007/s12517-022-10246-8

Fayaz, M., Nam, J., Dang, L. M., Song, H.-K., & Moon, H. (2024). Land-cover classification using deep learning with high-resolution remote-sensing imagery. *Applied Sciences*, 14(5). https://doi.org/10.3390/app14051844

Filik Iscen, C., Emiroglu, Ö., Ilhan, S., Arslan, N., Yilmaz, V., & Ahiska, S. (2008). Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat lake, Turkey. *Environmental Monitoring and Assessment*, 144(1), 269–276. https://doi.org/10.1007/s10661-007-9989-3

Li, C., Ma, Z., Wang, L., Yu, W., Tan, D., Gao, B., Feng, Q., Guo, H., & Zhao, Y. (2021). Improving the accuracy of land cover mapping by distributing training samples. *Remote Sensing*, 13(22). https://doi.org/10.3390/rs13224594

Liong, S. Y., & Sivapragasam, C. (2002). Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association*, 38(1), 173–186. https://doi.org/10.1111/j.1752-1688.2002.tb01544.x

Moumane, A., Bahouq, T., Karmaoui, A., Laghfiri, D., Yassine, M., Karkouri, J. Al, Batchi, M., Faouzi, M., Boulakhbar, M., & Youssef, A. A. (2025). Lake iriqui's remarkable revival: Field observations and a google earth engine analysis of its recovery after over half a century of desiccation. *Land*, 14(1). https://doi.org/10.3390/land14010104

Pang, H., Wang, X., Hou, R., You, W., Bian, Z., & Sang, G. (2023). Multiwater index synergistic monitoring of typical wetland water bodies in the arid regions of west-central Ningxia over 30 years. *Water*, 15(1). https://doi.org/10.3390/w15010020

Pham-Duc, B., Prigent, C., & Aires, F. (2017). Surface Water Monitoring within Cambodia and the Vietnamese Mekong delta over a year, with Sentinel-1 SAR observations. *Water*, 9(6). https://doi.org/10.3390/w9060366

Priyanka, N, S., Lal, S., Nalini, J., Reddy, C. S., & Dell'Acqua, F. (2023). DPPNet: An efficient and robust deep learning network for land cover segmentation from high-resolution satellite images. *IEEE Transactions on Emerging Topics in Computational Intelligence,* 7(1), 128–139. https://doi.org/10.1109/TETCI.2022.3182414

Šćepanović, S., Antropov, O., Laurila, P., Rauste, Y., Ignatenko, V., & Praks, J. (2021). Wide-area land cover mapping with Sentinel-1 imagery using deep learning semantic segmentation models. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 10357–10374. https://doi.org/10.1109/JSTARS.2021.3116094

Shabbir, A., Ali, N., Ahmed, J., Zafar, B., Rasheed, A., Sajid, M., Ahmed, A., & Dar, S. H. (2021). Satellite and scene image classification based on transfer learning and fine tuning of resnet50. *Mathematical Problems in Engineering*, 2021(1), 5843816. https://doi.org/https://doi.org/10.1155/2021/5843816

Solórzano, J. V, Mas, J. F., Gao, Y., & Gallardo-Cruz, J. A. (2021). Land use land cover classification with U-Net: Advantages of combining Sentinel-1 and Sentinel-2 imagery. *Remote Sensing*, 13(18). https://doi.org/10.3390/rs13183600

Uzun, M. (2024). Analysis of Manyas lake surface area and shoreline change over various periods with DSAS tool. *Turkish Journal of Remote Sensing,* 6(1), 35–56. https://doi.org/10.51489/tuzal.1443490

Xie, Y., Zeng, H., Yang, K., Yuan, Q., & Yang, C. (2023). Water-body detection in Sentinel-1 SAR images with DK-CO network. *Electronics*, 12(14). https://doi.org/10.3390/electronics12143163

Yao, P., Fan, H., & Wu, Q. (2025). Optimal drought index selection for soil moisture monitoring at multiple depths in China's agricultural regions. *Agriculture*, 15(4). https://doi.org/10.3390/agriculture15040423

Yilmaz, O. (2023). Monitoring water surfaces by remote sensing: The Case of Manyas Kusgolu, Ulubat, and Iznik lakes in Türkiye. *International Research in Engineering Sciences,* 52–66. https://doi.org/10.5281/zenodo.7744436

Zhang, W., Hu, B., & Brown, G. S. (2020). Automatic surface water mapping using polarimetric SAR data for long-term change detection. *Water*, 12(3). https://doi.org/10.3390/w12030872

Zhao, S., Tu, K., Ye, S., Tang, H., Hu, Y., & Xie, C. (2023). land use and land cover classification meets deep learning: A review. *Sensors*, 23(21). https://doi.org/10.3390/s23218966

Zhao, X., Zhang, J., Tian, J., Zhuo, L., & Zhang, J. (2020). Residual dense network based on channel-spatial attention for the scene classification of a high-resolution remote sensing image. *Remote Sensing*, 12(11). https://doi.org/10.3390/rs12111887