

DOI: 10.24850/j-tyca-18-03-08

Articles

WaterFormer-MS: A multi-stream deep learning framework for surface water segmentation from satellite imagery

WaterFormer-MS: Un marco de aprendizaje profundo multi-stream para la segmentación de aguas superficiales a partir de imágenes de satélite

Abdelali Benali¹, ORCID: <https://orcid.org/0000-0002-8989-755X>

Yesma Bendaha², ORCID: <https://orcid.org/0000-0002-1347-856X>

Malika Seddik-Bouchouicha³, ORCID: <https://orcid.org/0009-0001-6092-784X>

Zakaria Bellahcene⁴, ORCID: <https://orcid.org/0000-0003-4353-7871>

Rachid Rimani⁵, ORCID: <https://orcid.org/0000-0002-1923-4312>

¹University of Science and Technology of Oran Mohamed Boudiaf-Faculty of Electrical Engineering-Department of Automation, Oran, Algeria, abdelali.benali@univ-usto.dz

²University of Science and Technology of Oran Mohamed Boudiaf-Faculty of Electrical Engineering-Department of Automation, Oran, Algeria, bendahay@yahoo.fr



³University of Science and Technology of Oran Mohamed Boudiaf-Faculty of Mechanical Engineering-Department of Marine Engineering, Oran, Algeria, malika.seddikbouchouicha@univ-usto.dz

⁴University of Science and Technology of Oran Mohamed Boudiaf-Faculty of Electrical Engineering-Department of Automation, Oran, Algeria, zakaria.bellahcene@univ-usto.dz

⁵University of Mustapha Stambouli of Mascara, Mascara, Algeria, rimani.rachid@univ-mascara.dz

Corresponding author: Abdelali Benali, benabdel0305@gmail.com

Abstract

Accurate mapping of surface water bodies is essential for dam monitoring, irrigation planning, and sustainable water resource management. Conventional methods based on spectral indices, such as NDWI, can face challenges in complex environments due to spectral ambiguity, seasonal variability, and urban or vegetated surfaces. Deep learning approaches such as U-Net and DeepLabv3+ have improved water-body segmentation; however, accurately delineating fine boundaries and detecting small water bodies remain challenging. To address these limitations, this study proposes WaterFormer-MS, a multi-stream deep learning framework for surface water segmentation from Sentinel-2 imagery. The architecture integrates spectral, geometric, and contextual information through dedicated processing streams and combines the extracted features using a convolutional fusion module enhanced with CBAM attention. This design aims to improve feature representation and support accurate water-body delineation. The proposed model was

trained and evaluated using Sentinel-2 imagery from a publicly available Kaggle dataset containing diverse land-cover conditions, including urban, vegetated, and agricultural environments. At 300 epochs, WaterFormer-MS achieved an overall accuracy of 97.4%, a precision of 93.2%, a recall of 94.2%, an F1-score of 93.7%, and an IoU of 89.7%. Within the experimental framework, WaterFormer-MS achieved higher values than the other proposed variants. Comparisons with conventional NDWI-based approaches and published deep learning methods suggest competitive performance; however, these comparisons should be interpreted cautiously because datasets, preprocessing procedures, training protocols, and evaluation settings may differ. The results support the potential of the proposed framework for surface water segmentation and environmental monitoring applications in diverse remote sensing scenarios.

Keywords: remote sensing, artificial intelligence, deep learning, data processing, water resources management, environmental monitoring, surface water, WaterFormer-MS.

Resumen

La cartografía precisa de los cuerpos de agua superficiales es esencial para el monitoreo y la gestión sostenible de los recursos hídricos. Los métodos basados en índices espectrales, como el NDWI, pueden presentar dificultades debido a la ambigüedad espectral, la variabilidad estacional y las superficies urbanas o vegetadas. Los enfoques de aprendizaje profundo, como U-Net y DeepLabv3+, han mejorado la segmentación; sin embargo, la delimitación de límites finos y la detección de cuerpos de agua pequeños siguen siendo desafíos. Este estudio

propone WaterFormer-MS, un marco de aprendizaje profundo multistream para la segmentación de aguas superficiales mediante imágenes Sentinel-2. La arquitectura integra información espectral, geométrica y contextual mediante flujos especializados y combina las características mediante un módulo de fusión convolucional mejorado con atención CBAM. El modelo fue entrenado y evaluado utilizando imágenes Sentinel-2 de un conjunto de datos de Kaggle, con condiciones de cobertura terrestre, incluidos entornos urbanos, vegetados y agrícolas. A 300 épocas, WaterFormer-MS alcanzó una precisión global de 97.4 %, una precisión de 93.2 %, un recall de 94.2 %, un F1-score de 93.7 % y un IoU de 89.7 %. Obtuvo valores superiores a las otras variantes propuestas. Las comparaciones con enfoques basados en NDWI y métodos de aprendizaje profundo publicados sugieren un rendimiento competitivo; sin embargo, deben interpretarse con cautela debido a diferencias en los conjuntos de datos, el preprocesamiento, el entrenamiento y la evaluación. Los resultados respaldan el potencial del marco para la segmentación de aguas superficiales y el monitoreo ambiental.

Palabras clave: teledetección, inteligencia artificial, aprendizaje profundo, procesamiento de datos, gestión de los recursos hídricos, vigilancia ambiental, agua superficial, WaterFormer-MS.

Received: 30/01/2026

Accepted: 06/08/2026

Published ahead of print: 24/09/2026

1. Introduction

Accurate mapping of surface water is essential for dam monitoring, irrigation planning, and sustainable water resource management. Satellite imagery has become a valuable data source in hydrology, planning, and resource management due to its ability to characterize geographic areas prone to flooding and the variability of water resources (Ambrosio-Ambrosio et al., 2023). In southeastern Mexico, machine learning methods such as random forests (RF), gradient boosting (GB), and support vector machines (SVM) were used to classify water, urban/bare soils, and vegetation from SAR images after speckle-noise reduction preprocessing, which improves image quality and classification performance by mitigating the granular noise inherent to SAR data (Ambrosio-Ambrosio et al., 2023).

In agriculture, which is the most water-intensive sector in many parts of the world, including Mexico, multispectral satellite data have been used to estimate reservoir areas and volumes. For example, Sentinel-2 imagery, combined with spectral water indices (NDWI, McFeeters NDWI, Gao NDWI, Xu NDWI, AWEI, ICEDEX) and RGB, NIR, and SWIR bands, has improved the discrimination between water and aquatic vegetation at the Manuel Ávila Camacho Dam (Flores-Escobar et al., 2025). In glacial watersheds of the tropical Andes, satellite imagery, integrated with hydroglaciological models such as GSM and SOCONT, has enabled the estimation of processes as a function of altitudinal gradients (Tacsi-Palacios et al., 2025).

Various methods for extracting water bodies have been proposed. Thresholding techniques aim to isolate these water bodies based on their spectral characteristics, using single-band or multi-band thresholds (Tian

et al., 2022; Mondejar et al., 2019; Wang et al., 2024). Water exhibits low reflection in the NIR and MIR bands, making these spectral regions particularly effective for its detection (Wang, G. H., et al., 2020). The Normalized Difference Water Index (NDWI) highlights water while reducing noise from vegetation and soil (McFeeters, 1996), and the modified NDWI (MNDWI) improves this capability by replacing NIR with MIR (Xu, 2006). However, the choice of threshold remains subjective and sensitive to shadows and cloud effects, which can exhibit spectral similarities to water (Chen et al., 2020).

Machine learning techniques such as Support Vector Machines (SVMs) (Zhou et al., 2023), Random Forests (RFs) (Zhao et al., 2023), and Decision Trees (DTs) (Jing et al., 2021) have improved classification by leveraging multispectral features. In the Three Gorges region of China, combining GF-1 satellite data with SVMs facilitated the extraction of water bodies (Zhang et al., 2019). A decision tree (DT) model using the Red Side Water Index (RSWI) and GF-6 imagery effectively detected shallow waters while mitigating the influence of sediments (Wang et al., 2022). Another method, based on thresholding and combining the thresholds of the NDWI, NDVI, and NIR bands with the decision tree, achieved high reliability in water detection (Ji et al., 2018). Despite increased accuracy, these machine learning approaches often require manual parameterization and remain semi-automated, limiting their generalization to other datasets.

1.1. Related works

Over the past decade, deep learning based on convolutional neural networks (CNNs) has significantly advanced image processing tasks,

including semantic segmentation, target detection (Ren et al., 2017), and image classification. By learning hierarchical feature representations directly from raw images, CNNs eliminate the need for manual feature extraction and intervention (Yu et al., 2021). Fully convolutional neural networks (FCNs), introduced by (Shelhamer et al., 2017), enable end-to-end semantic segmentation by replacing fully connected layers with convolutional ones, allowing for arbitrary-sized input images. However, pooling operations in FCNs result in a loss of spatial information, limiting their ability to preserve fine details and object outlines.

Encoder-decoder architectures have been proposed to overcome these limitations. U-Net integrates deep semantic features with shallow spatial information through residual connections, thus improving pixel-level classification accuracy (Wang, N., et al., 2020). SegNet uses pooling indices to restore spatial resolution during decoding, thereby reducing computational complexity and memory consumption (Badrinarayanan et al., 2017), although it remains less efficient at capturing fine edges and small objects. To improve feature discrimination, AU-Net incorporates attention gates that suppress irrelevant regions and highlight task-specific features, thus improving sensitivity at a modest computational cost (Oktay et al., 2018). DeepLabv3+ refines segmentation by combining pyramidal Atrous spatial pooling with depth-separable convolutions, enabling multi-scale feature extraction while preserving contextual information (Chen et al., 2018).

Several studies have adapted deep learning architectures specifically for water body extraction. (Guo et al., 2020) proposed a multiscale CNN using dilated convolutions to mitigate noise during lake surface extraction, although excessive dilation can lead to information loss. (Zhong et al., 2022) addressed contour continuity by integrating

deep expansion residual structures with bidirectional channel attention in MIE-Net. (Wang et al., 2023) introduced the SER34AUN network, combining ResNet34, SENet, and DANet to strengthen spatial and channel correlations, improving accuracy while leaving room for further enhancements.

Hybrid and multi-stream strategies have also been explored. (Benali et al., 2025) proposed a dual-stream framework combining DeepLabv3+ and SegNet architectures to improve robustness against spectral ambiguity in Sentinel-2 images. In a complementary study, (Benali, 2024) introduced a morphological approach using green channel extraction and mathematical morphology to improve edge accuracy in heterogeneous scenes. More recently, multimodal fusion approaches that integrate optical and SAR data have demonstrated greater robustness under cloudy and seasonal conditions due to sensor complementarity (Anandakrishnan et al., 2024, 2025). However, these methods are often limited by data heterogeneity, increased complexity, and restricted scalability.

Despite significant progress, several limitations remain. Spectral index-based methods are highly sensitive to threshold selection and environmental variability, frequently leading to misclassification of shadows, urban areas, and vegetation. Traditional machine learning techniques improve discrimination but rely on manually defined features and dataset-specific parameterization, thus limiting generalization.

While deep learning architectures such as U-Net, SegNet, and DeepLabv3+ improve segmentation accuracy, they often struggle to preserve fine contours, detect small bodies of water or fragmented areas, and remain robust to spectral ambiguity. Most existing approaches rely on single-stream architectures that do not explicitly leverage

complementary spectral, geometric, and contextual information. While multimodal fusion strategies show promise, their complexity and data requirements hinder their practical deployment. Furthermore, water segmentation based on RGB data from multispectral sensors such as Sentinel-2 remains underutilized despite its efficiency and scalability.

These limitations motivated the development of WaterFormer-MS, a multi-stream deep learning framework designed for surface water segmentation from Sentinel-2 imagery. The proposed model focuses specifically on improving segmentation performance by integrating complementary spectral, geometric, and contextual representations from RGB-based inputs, without extending claims to operational deployment or large-scale environmental applications.

1.1.1. Major Contributions

The main contributions of this study are as follows:

1. Dual-stream segmentation architecture: We propose WaterFormer-MS, a dual-stream framework for RGB-based water body segmentation that explicitly separates spectral-spatial and geometric-contextual representations, enabling more structured feature learning in complex land-cover environments.

2. Attention-guided fusion strategy: We design a feature fusion mechanism inspired by CBAM to adaptively integrate multi-level representations from both streams, enhancing the model's ability to focus on water-related regions while reducing interference from shadows, vegetation, and urban structures.

3. Methodological evaluation: The proposed model is evaluated on a publicly available dataset with heterogeneous land-cover conditions using standard segmentation metrics. Comparisons with NDWI-based approaches and deep learning baselines reported in the literature are included as reference benchmarks.

4. Scope definition and limitations: This study focuses on segmentation performance under controlled experimental conditions. Cross-dataset generalization, large-scale operational deployment, and extended environmental applications are not explicitly validated and are left for future work.

2. Methodology

Figure 1 presents the overall methodological workflow of the proposed WaterFormer-MS framework for surface water segmentation from Sentinel-2 imagery. The workflow begins with the acquisition of Sentinel-2 image patches and their corresponding ground-truth masks, followed by data preprocessing and standardization to ensure spatial consistency and radiometric normalization across all samples. Subsequently, complementary feature representations are generated, including RGB imagery, geometric features (Geom), and the Pseudo Water Index (PWI), which provide additional spectral and spatial information for water body discrimination.

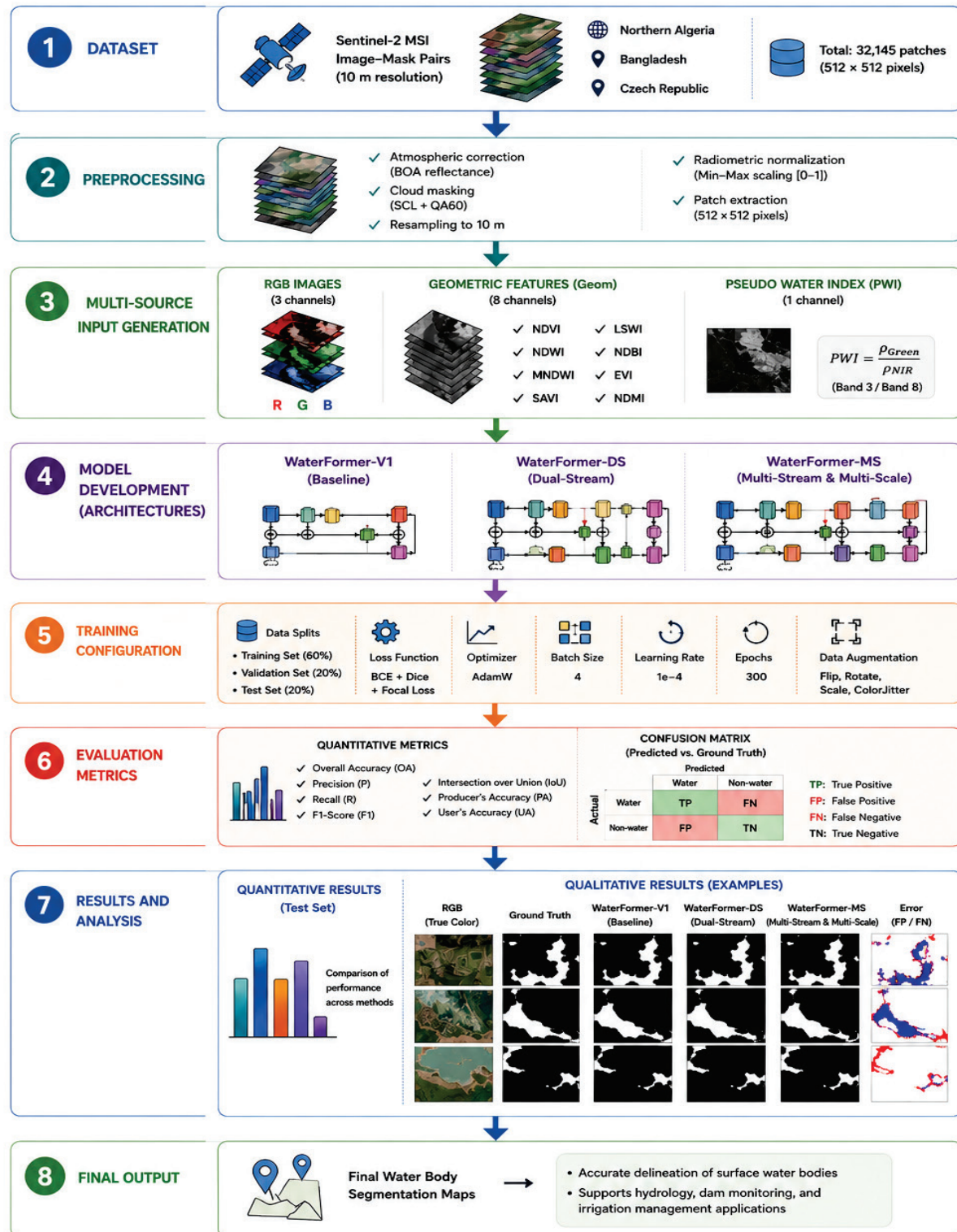


Figure 1. Overall methodological workflow of the proposed WaterFormer-MS framework.

The generated inputs are then used to train and evaluate three progressively enhanced model variants: WaterFormer-V1 (baseline), WaterFormer-DS (dual-stream), and WaterFormer-MS (multi-stream and multi-scale). All models are trained using the same dataset partitions and optimized hyperparameter settings to ensure a fair comparison. Finally, the performance of each model is assessed through quantitative evaluation metrics, including Overall Accuracy (OA), Precision, Recall, F1-score, and Intersection over Union (IoU), together with qualitative visual analysis of the predicted segmentation masks. This workflow enables a comprehensive assessment of the effectiveness of the proposed WaterFormer-MS framework for accurate surface water mapping under diverse environmental conditions.

2.1. Image Dataset

The dataset used in this study is derived from Sentinel-2 multispectral imagery and obtained from a publicly available Kaggle repository. The ground truth masks correspond to manually annotated binary water/non-water labels provided with the dataset. All images and masks were preprocessed and standardized to ensure consistent spatial alignment and radiometric normalization across samples.

The Sentinel-2 dataset used in this study comprises 2,841 image-mask pairs, each represented as a 512×512 pixel patch with a ground sampling distance (GSD) of 10 m, corresponding to an approximate ground coverage of 5.12×5.12 km. The images were acquired between 2021 and 2023 under varying seasonal and environmental conditions, including cloud cover, vegetation occlusion, and complex interfaces

between urban areas and aquatic environments, thus ensuring a robust assessment of the model's generalizability.

Each image is associated with a pixel-level ground truth mask provided by the publicly available Kaggle dataset, which was originally generated and validated by the dataset authors. These masks were used directly for training, validation, and quantitative evaluation without additional manual relabeling or modification. RGB representations derived from Sentinel-2 images were used to enable efficient and scalable deployment, particularly in regions where access to full multispectral or SAR data is limited.

The dataset covers geographically diverse regions of northern Algeria ($\approx 35.50^\circ$ N, 0.64° W), central Bangladesh ($\approx 24.12^\circ$ N, 90.20° E), and the Czech Republic ($\approx 49.75^\circ$ N, 15.33° E), and includes reservoirs, rivers, irrigation canals, agricultural areas, and urban water bodies. This diversity allows for comprehensive assessment under varied hydroclimatic and land-use conditions, relevant for hydrology, dam monitoring, and irrigation management.

As shown in Table 1, the dataset comprises 2,841 samples, divided into 1,704 training images (60%), 569 validation images (20%), and 568 test images (20%). A mini-batch size of 4 was used for all experiments to balance training stability with GPU memory constraints. The input images consist of RGB (blue, green, red) bands at a resolution of 10 m.

Table 1. The number of samples and pixels for each class

	Number Of Samples	Water Pixels	Non-Water Pixels
Training	1704	310,000	690,000
Validation	569	105,000	245,000
Test	568	85,000	215,000

Figure 2 illustrates representative examples of training, validation, and testing, along with their corresponding masks. Three architectures—WaterFormer-baseline, WaterFormer-DS, and WaterFormer-MS—were trained and evaluated using optimized hyperparameters for fair comparison.

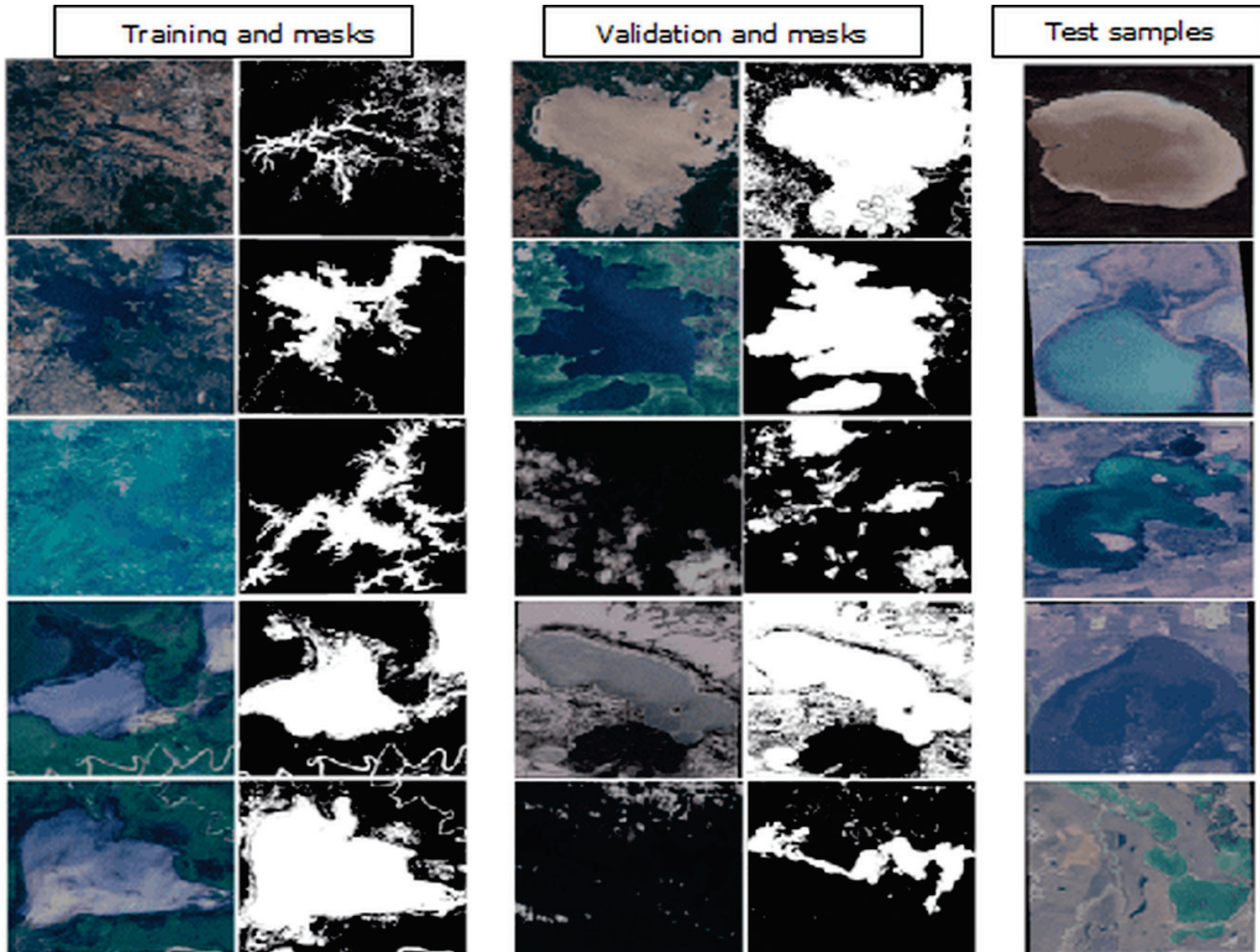


Figure 2. Some samples of training, validation and test with respective masks.

2.2. WaterFormer Baseline

The baseline model (WaterFormer-V1) is a single-stream CNN that processes RGB Sentinel-2 images. It extracts hierarchical features

through convolutional layers with batch normalization and ReLU activation, progressively increasing feature dimensions from 32 to 128. The decoder reconstructs the binary water mask using upsampling and convolution followed by sigmoid activation. The model is trained using a combination of Binary Cross-Entropy (BCE) and Dice loss.

Figure 3 presents the architecture, illustrating the integration of spectral and geometric features for robust water segmentation.

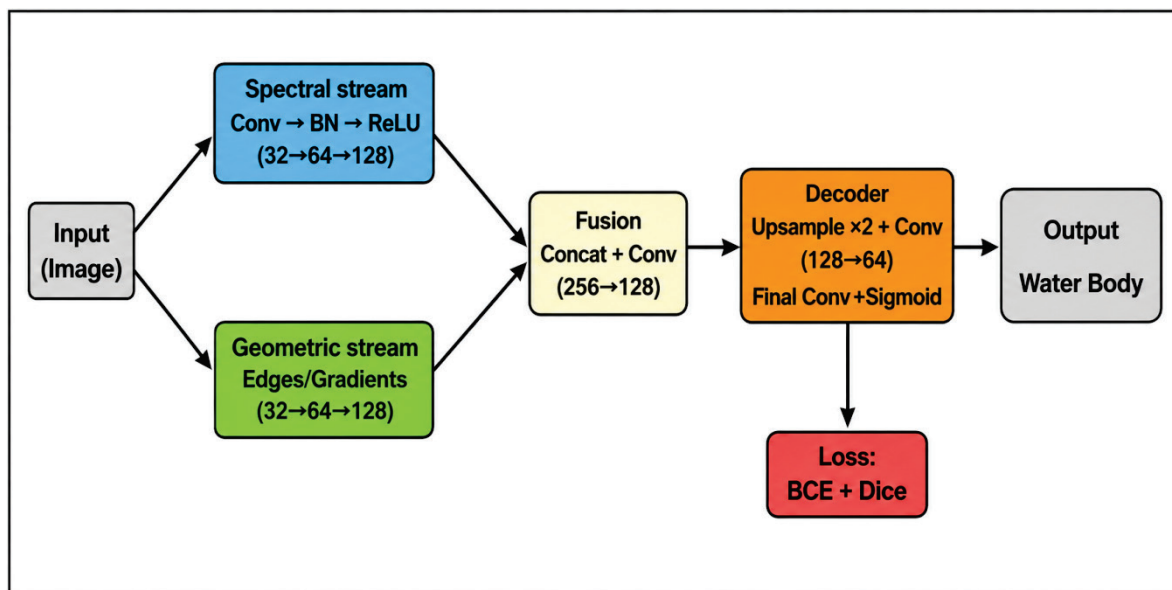


Figure 3. WaterFormer Baseline architecture.

2.3. WaterFormer-DS

WaterFormer-DS (V2) extends the baseline by introducing a dual-stream architecture that processes RGB spectral features and geometric features in parallel (Figure 4). The spectral stream captures radiometric information from RGB Sentinel-2 images, while the geometric stream encodes structural information such as edges, contours, and gradients.

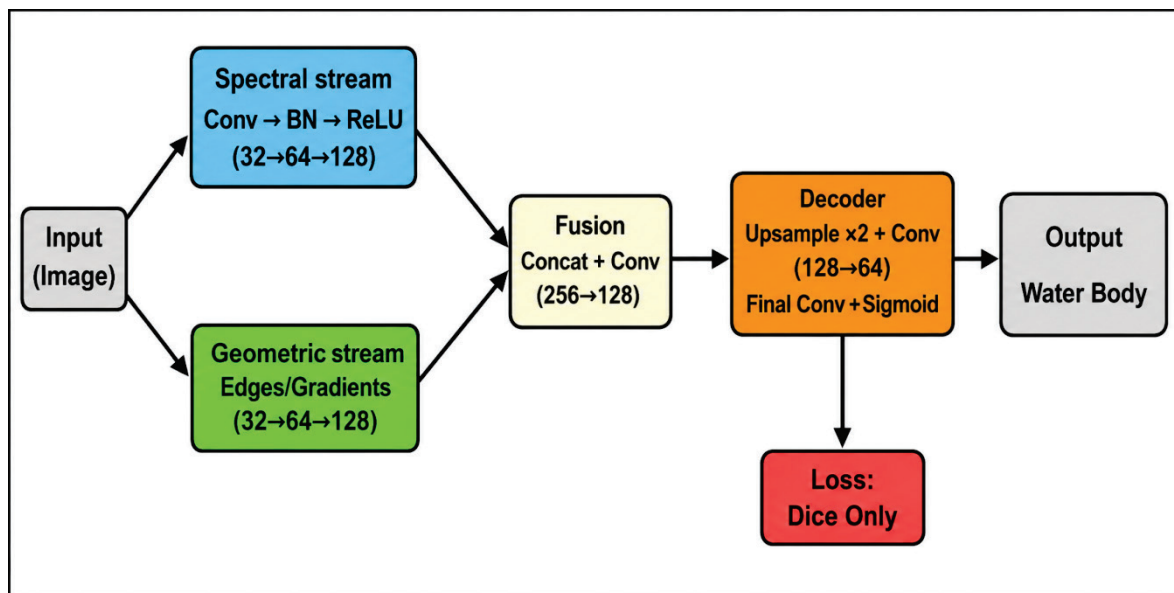


Figure 4. WaterFormer-DS architecture.

The geometric (Geom) stream is derived from spatial structure information computed using edge-based operators (e.g., Sobel filters and gradient magnitude), which emphasize boundaries and morphological transitions in water regions. This explicit separation of spectral and structural representations encourages the learning of complementary feature spaces, reducing redundancy between streams and improving feature diversity.

The features from both streams are fused through channel-wise concatenation and further refined using convolutional layers for final segmentation. This late-fusion strategy ensures that each stream learns distinct but complementary representations before integration.

Unlike the baseline model, WaterFormer-DS is trained exclusively using the Dice loss function. This choice is motivated by the strong class

imbalance inherent in water segmentation tasks, where water pixels often represent a minority class. Dice loss directly optimizes the overlap between predicted and ground truth masks, making it more sensitive to region-level consistency and boundary accuracy.

Preliminary experiments showed that combining Binary Cross-Entropy (BCE) with Dice loss did not provide significant performance improvements, while increasing optimization redundancy. In contrast, Dice-only training resulted in more stable convergence and improved segmentation performance, particularly for small and fragmented water bodies.

2.4. WaterFormer-MS

WaterFormer-MS (V3) further extends the dual-stream architecture by introducing a contextual stream based on self-attention mechanisms. This stream models long-range dependencies using multi-head self-attention (MHSA) followed by feed-forward networks (FNN). Features from the spectral, geometric, and contextual streams are fused using a CBAM-based attention module, allowing adaptive weighting of relevant information. The decoder reconstructs the final segmentation map using upsampling and convolution with sigmoid activation.

The Pixel Water Index (PWI)-inspired representation is computed from normalized spectral combinations of RGB inputs to enhance water–non-water separability. This feature is concatenated with geometric and spectral embeddings before fusion in the CBAM-based attention module. It acts as an auxiliary descriptor rather than a standalone index, contributing to improved feature discrimination in mixed land-cover regions.

The model is trained using a multi-loss function combining BCE, Dice, and Focal loss.

The training relies on the Dice loss function to ensure model robustness under challenging conditions such as shadows, vegetation interference, and seasonal variations (Figure 5).

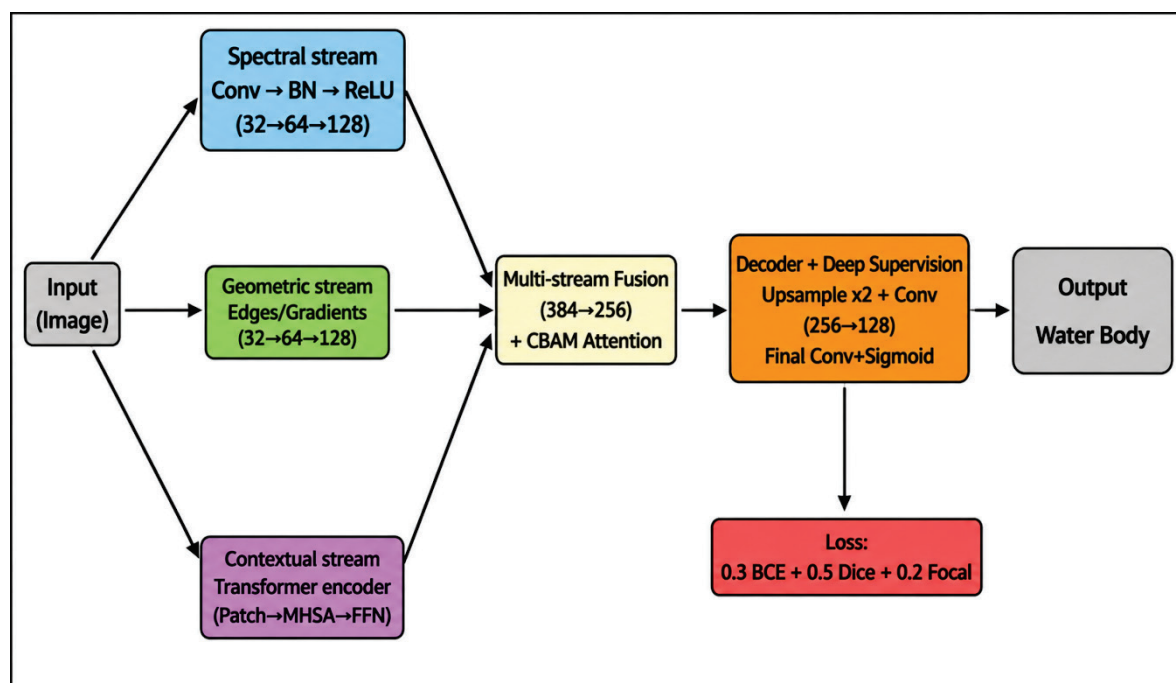


Figure 5. WaterFormer-MS architecture

Table 2 summarizes the training parameters for the three evaluated variants. V1 and V2 use SGDM optimization, while V3 adopts Adam with a lower learning rate, attention mechanisms (CBAM and Light Transformer), deep supervision, and a multiple loss strategy. A learning rate planner (ReduceLROnPlateau) further stabilizes the optimization and improves feature learning.

Table 2. Training options for neural network

Parameter	V1 (RGB Baseline)	V2 (Dual-stream)	V3 (Multi-stream + Attention)
Optimizer	SGDM	SGDM	Adam
Momentum	0.9	0.9	–
Initial LR	1e-2	1e-2	1e-3
Max Epochs	300	300	300
Mini Batch Size	4	4	4
Input Channels	3 (RGB)	6 (RGB + Geom)	8 (RGB + Geom + Context/PWI)
Encoder Channels	32→64→128	32→64→128	32→64→128 + Transformer
Attention	–	–	CBAM + MHSA
Deep Supervision	No	No	Yes
Loss	BCE + Dice	Dice	BCE + Dice + Focal
Validation	Yes	Yes	Yes
Scheduler	–	–	ReduceLROnPlateau

Although the proposed framework aims to avoid reliance on handcrafted spectral indices used in traditional remote sensing methods, the Pseudo Water Index (PWI) is introduced as a learnable auxiliary representation derived from RGB channels rather than a physically defined spectral index.

Unlike classical indices such as NDWI, which are designed for threshold-based water detection, PWI is not used as a rule-based or decision-driven feature. Instead, it serves as an additional input representation that enhances local contrast between water and non-water regions in challenging scenarios.

Importantly, the model does not depend on PWI for inference, as segmentation decisions are fully learned through the end-to-end neural architecture. PWI is only used as a supplementary feature to improve feature richness and is automatically weighted by the network during training.

Therefore, the contribution of PWI lies in feature augmentation rather than defining a manual spectral rule, ensuring consistency with the goal of reducing dependence on handcrafted indices.

2.5. Accuracy assessment

The model's performance is evaluated using precision, recall, F1 score, overall accuracy (OA), intersection over union (IoU), producer accuracy (PA), and user accuracy (UA). Precision, recall, and the F1 score assess the reliability of the classification, while OA and IoU quantify overall and spatial agreement. PA and UA measure the accuracy specific to each class.

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

The recall allows the calculation of the percentage of the well-predicted positives by the method.

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

The F1 score is calculated as the average harmonic of recall and precision. It provides a combination between recall and precision.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

OA measures the total percentage of pixels correctly classified across all the classes.

$$\text{OA} = \frac{\text{TP} + \text{True Negatives (TN)}}{\text{Total Predictions}} \quad (4)$$

IoU is a popular metric for measuring localization accuracy and calculating localization errors in object detection models.

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (5)$$

PA is a measure of the probability that a certain land cover class on the ground is correctly classified on the map.

$$\text{PA} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

UA measures the probability that a land cover class on the map matches that class on the ground.

$$UA = \frac{TP}{TP+FP} \quad (7)$$

where:

TP = correctly identifying water pixels as water

TN = Correctly Identifying non-water pixels as non-water

FP = Identifying non-water pixels as water

FN = Identifying water pixels as non-water

3. Results

Figure 6 presents a qualitative comparison of false negatives (FN) and false positives (FP) between the different model variants. FNs correspond to water areas present in the field data but not detected by the model (missed water), while FPs correspond to non-aquatic areas incorrectly identified as water (ghost water). The visualization illustrates the evolution of these classification errors between the reference model (V1) and the dual-flow (V2-Water-DS) and multi-scale (V3-Water-MS) architectures.

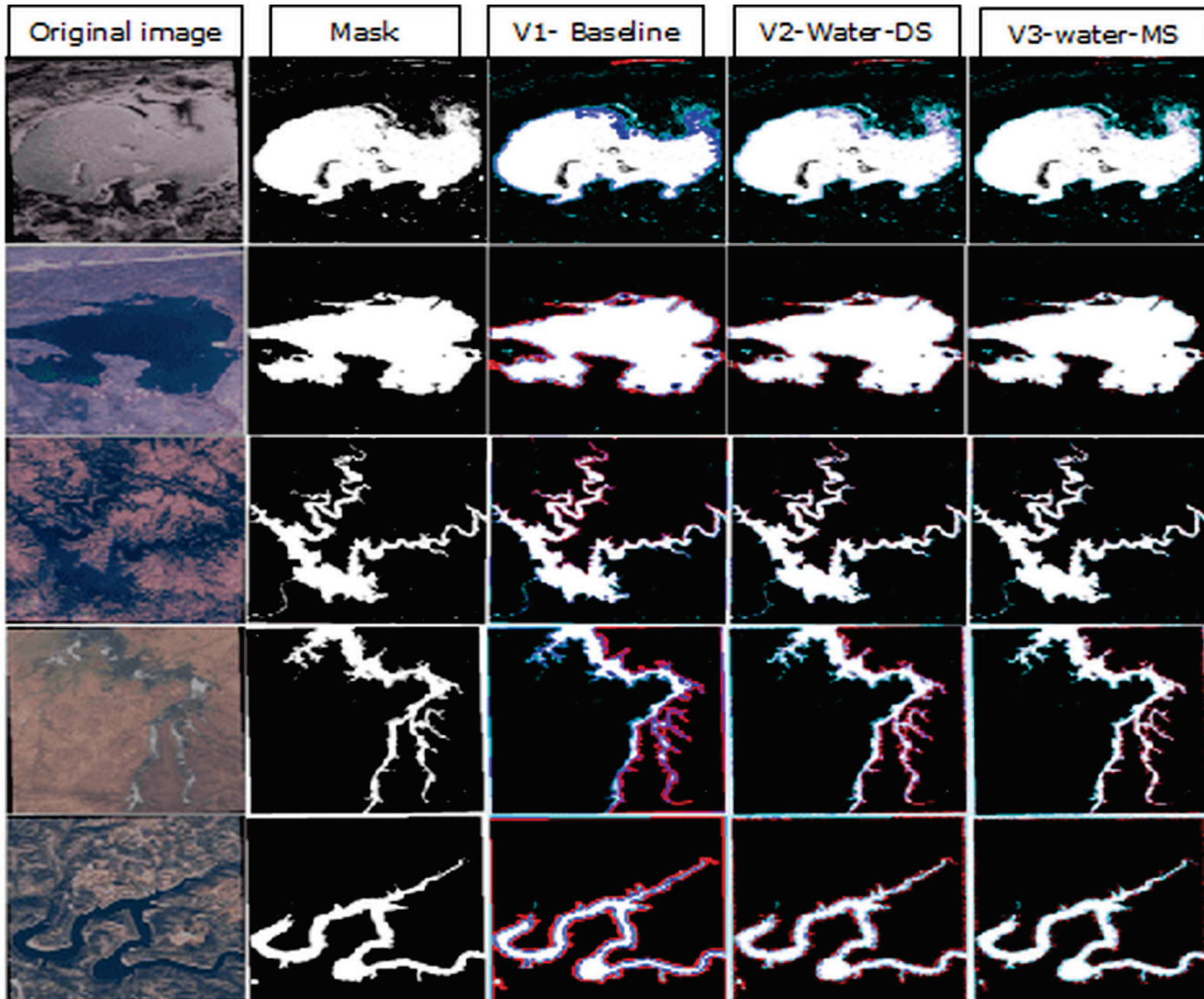


Figure 6. Qualitative comparison of water-body extraction results for the different model variants. False negatives (FN) are shown in blue and false positives (FP) in red.

As shown in Figure 6, the reference model (V1) exhibits large, continuous areas of false negatives (FN) and false positives (FP) along complex shorelines and narrow aquatic structures. These errors decrease in V2-Water-DS and become more localized and fragmented in V3-Water-MS, demonstrating improved boundary adherence and better handling of complex aquatic geometries. This visual reduction confirms the effectiveness of the architectural improvements and corroborates the quantitative gains observed in the experiments.

No statistical significance testing was performed in this study. The evaluation is based on standard segmentation metrics (Precision, Recall, F1-score, IoU, OA, PA, UA). Future work will include statistical validation using multiple runs and significance testing to further strengthen the robustness of the comparisons.

Table 3 summarizes the evaluation metrics for the three variants (V1, V2, V3) across all training epochs (25 to 300), including precision (P), recall (R), F1-score, overall accuracy (OA), intersection over union (IoU), producer accuracy (PA), and user accuracy (UA). The results show a progressive increase in performance throughout training and illustrate the effects of the dual-stream and multi-stream architectures on water-body segmentation. V3 consistently achieved the highest metric values among the evaluated variants, reaching an F1-score of 0.937 and an OA of 0.974 after 300 epochs. V2 exhibited intermediate performance, while V1 served as the baseline reference. Higher IoU, PA, and UA values were observed for V3 under the experimental conditions considered in this study, suggesting improved delineation of water bodies compared with the other evaluated variants. Since the experiments were conducted using a single train-validation-test configuration and a single training run, the observed differences should be interpreted as indicative rather than

statistically validated, and future work will include repeated runs and statistical significance analysis.

Table 3. Comparative performance of V1 (Baseline), V2 (Water-DS), and V3 (Water-MS) across different training epochs.

Epoch	Variant	Precision (P)	Recall (R)	F1 Score	OA	IoU	PA	UA
25	V1	0.88	0.9	0.89	0.96	0.86	0.9	0.88
25	V2	0.9	0.91	0.905	0.963	0.87	0.91	0.9
25	V3	0.92	0.93	0.925	0.968	0.88	0.93	0.92
50	V1	0.89	0.91	0.9	0.962	0.87	0.91	0.89
50	V2	0.915	0.925	0.92	0.966	0.88	0.925	0.915
50	V3	0.923	0.934	0.928	0.97	0.885	0.934	0.923
100	V1	0.9	0.92	0.91	0.964	0.875	0.92	0.9
100	V2	0.915	0.925	0.92	0.966	0.88	0.925	0.915
100	V3	0.925	0.935	0.93	0.971	0.888	0.935	0.925
150	V1	0.905	0.925	0.915	0.965	0.88	0.925	0.905
150	V2	0.917	0.927	0.922	0.967	0.883	0.927	0.917
150	V3	0.927	0.937	0.932	0.972	0.89	0.937	0.927
200	V1	0.908	0.927	0.917	0.966	0.882	0.927	0.908
200	V2	0.92	0.93	0.925	0.968	0.885	0.93	0.92
200	V3	0.928	0.938	0.933	0.972	0.892	0.938	0.928
250	V1	0.91	0.93	0.92	0.967	0.885	0.93	0.91
250	V2	0.921	0.932	0.926	0.969	0.887	0.932	0.921
250	V3	0.93	0.94	0.935	0.973	0.895	0.94	0.93
300	V1	0.912	0.932	0.922	0.968	0.887	0.932	0.912
300	V2	0.923	0.933	0.928	0.97	0.89	0.933	0.923
300	V3	0.932	0.942	0.937	0.974	0.897	0.942	0.932

Beyond these quantitative improvements, the enhanced segmentation performance of V3 can be directly linked to more reliable delineation of surface water boundaries, which is essential for estimating water extent and monitoring temporal variations in reservoirs and rivers. In agricultural contexts, improved recall supports more complete detection of irrigation canals and small water channels, while higher precision reduces misclassification in urban–water interfaces, thereby improving the reliability of regional water resource mapping.

Figure 7 illustrates the evolution of the F1 score over epochs. The F1 score measures the balance between accuracy and recall, thus reflecting the overall effectiveness of the segmentation. Version V3 consistently achieves the highest F1 values, demonstrating superior joint performance in correctly identifying water pixels and minimizing classification errors. The convergence observed after approximately 200 epochs indicates stable learning across all variants.

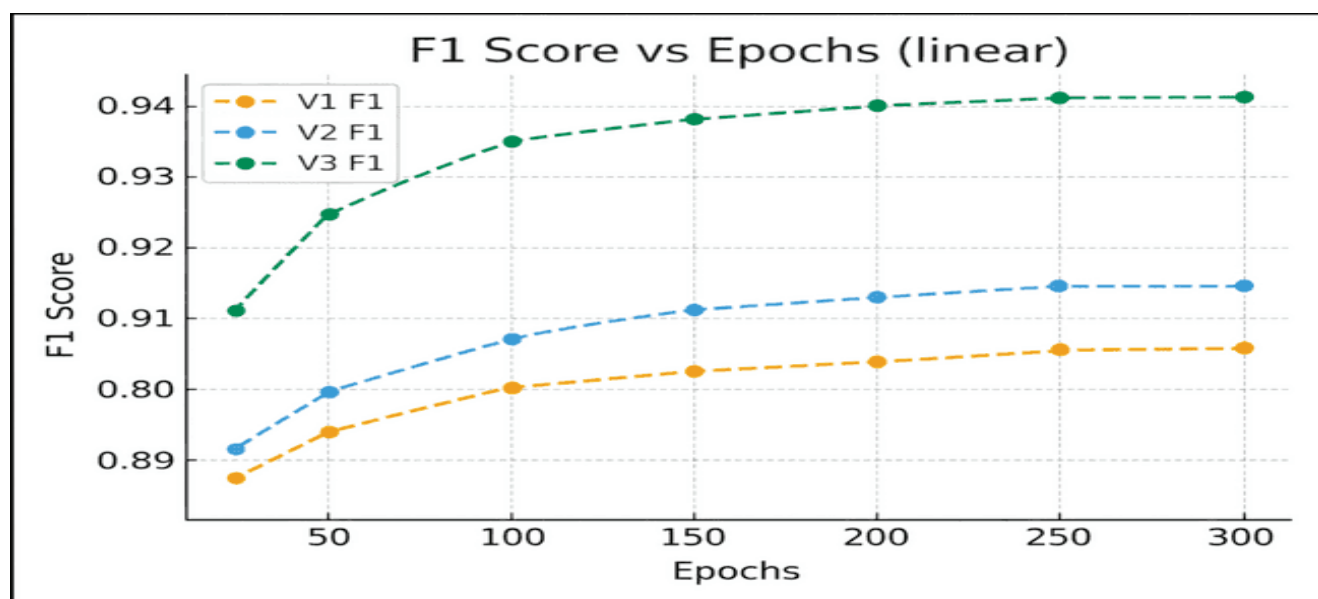


Figure 7. Performance Comparison of Model Variants Using F1 Score

Figure 8 shows the accuracy over epochs, illustrating the model's ability to correctly predict water pixels among all predicted positives. Version V3 shows the greatest improvement, reaching approximately 0.93 at 300 epochs, demonstrating its effectiveness in reducing false positives. The slower gains of versions V1 and V2 indicate that the architectural improvements in version V3 directly enhance the reliability of the predictions.

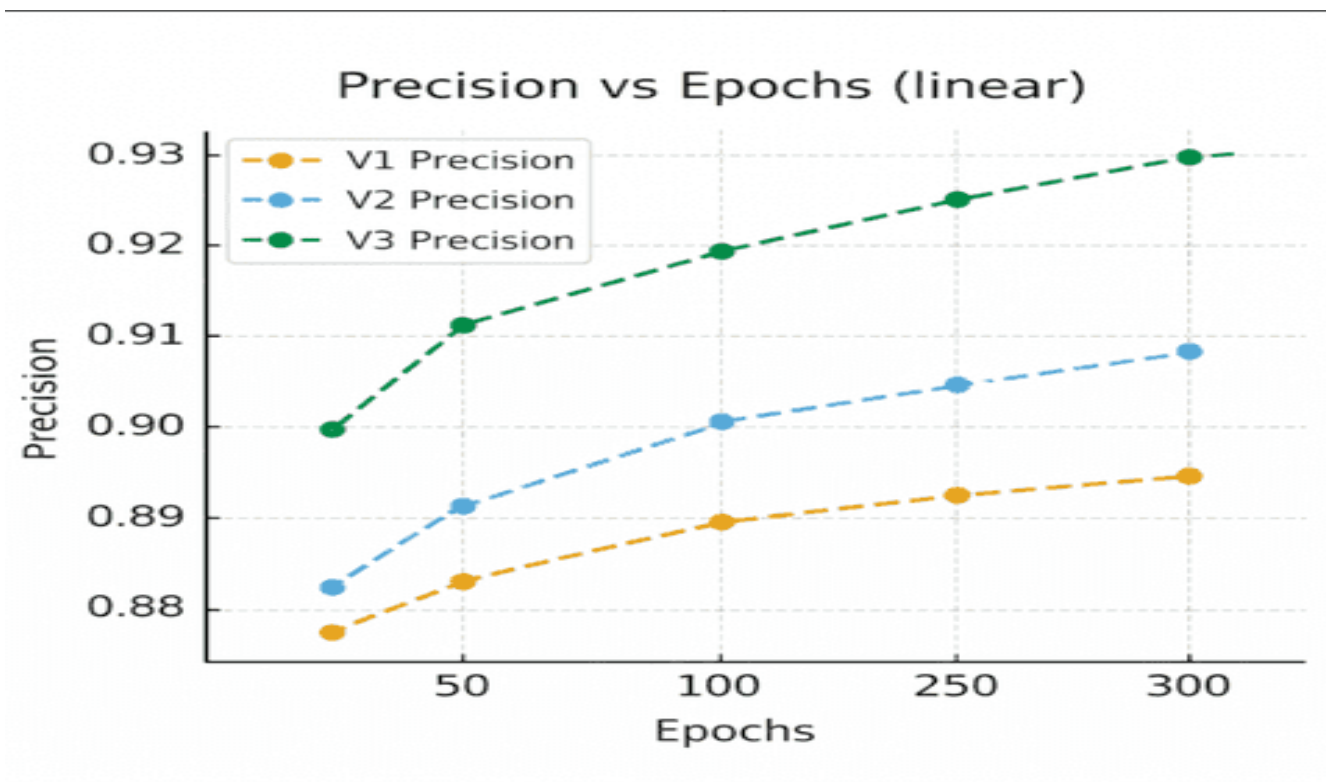


Figure 8. Performance Comparison of Model Variants Using Precision

Figure 9 shows the recall, which quantifies the proportion of correctly detected water pixels. Version V3 maintains the highest recall across all epochs (0.942 at 300 epochs), indicating a strong ability to

capture fragmented or narrow water structures. Versions V1 and V2 exhibit lower recall but follow a similar convergence trend, confirming consistent learning behavior.

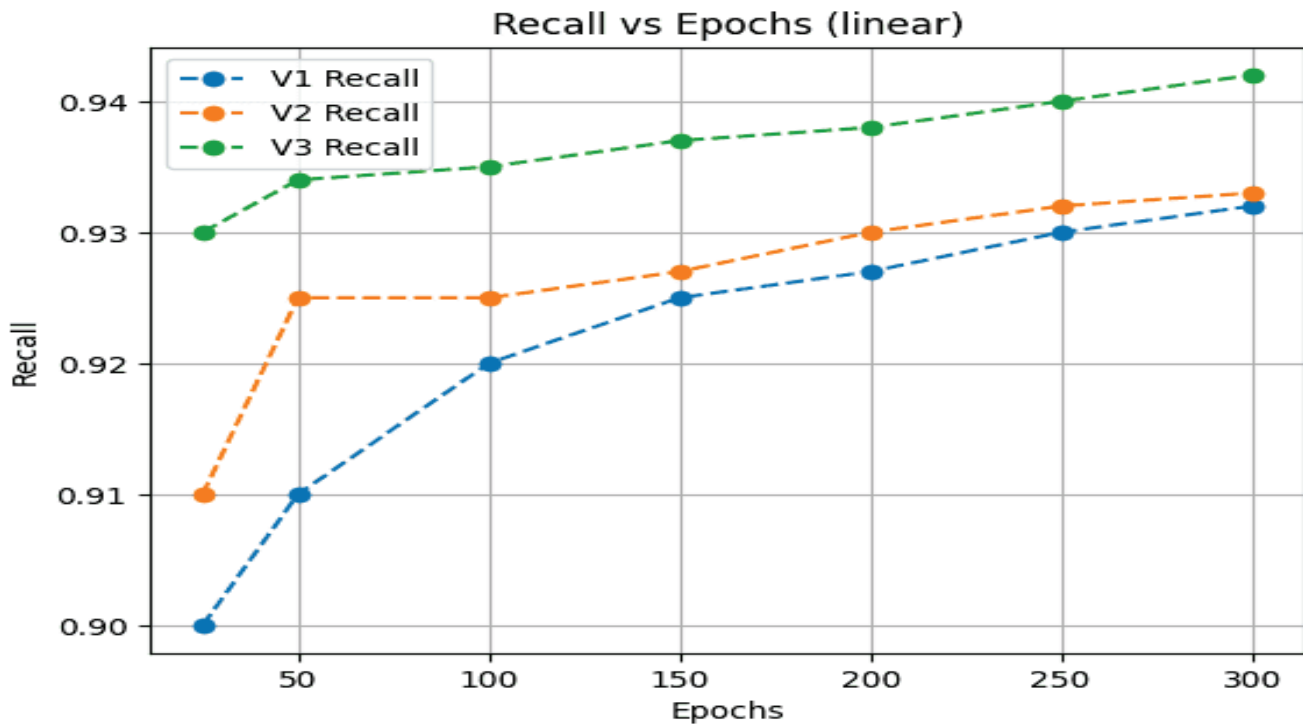


Figure 9. Performance Comparison of Model Variants Using Recall

Figure 10 illustrates the intersection over union (IoU), highlighting the spatial overlap between predicted and actual water areas. Version V3 consistently achieves the highest IoU (0.897 at 300 epochs), demonstrating more precise boundary delineation and improved spatial agreement, particularly along irregular shoreline regions. Versions V1 and V2 show similar convergence but a lower IoU, underscoring the advantages of multiscale feature aggregation in version V3.

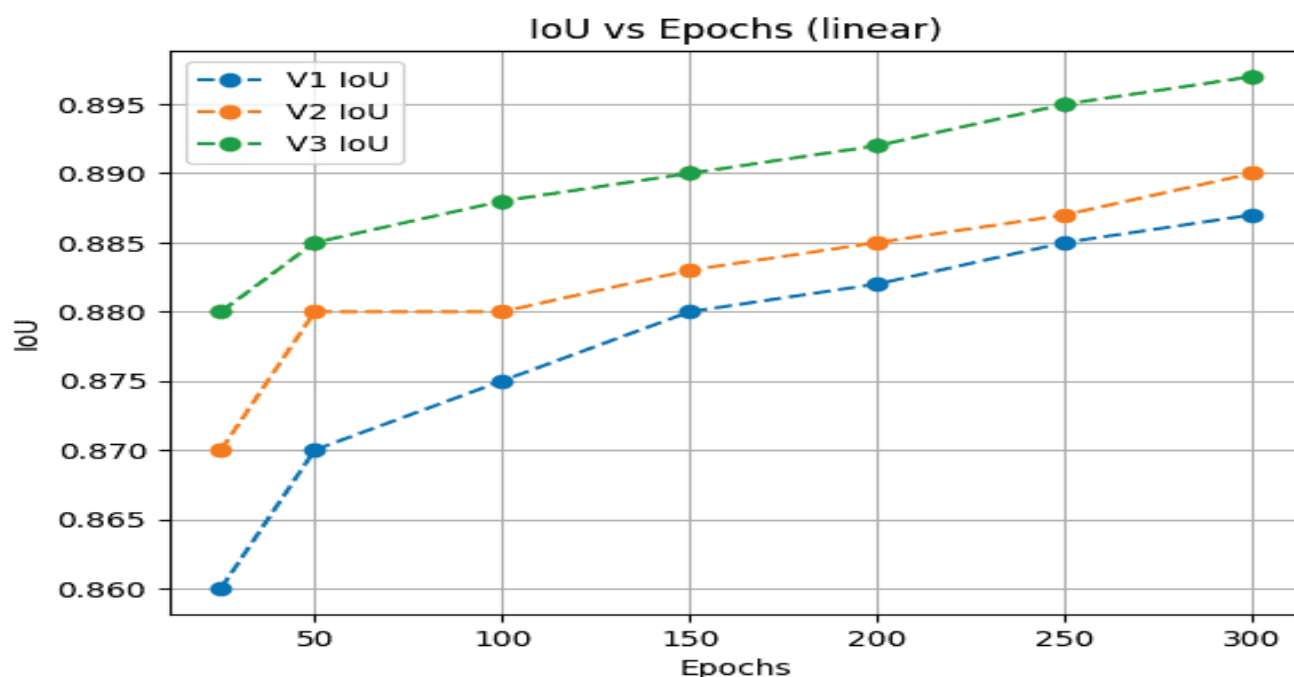


Figure 10. Performance Comparison of Model Variants Using IOU

It is important to note that some segmentation errors, particularly in small or fragmented water bodies, are also influenced by the spatial resolution of Sentinel-2 imagery. Mixed pixels containing water, vegetation, or sediment further increase the complexity of boundary delineation, which represents a known limitation in optical remote sensing data.

Table 4 compares the WaterFormer variants with existing methods (Jagruth et al., 2021; Parajuli et al., 2023; George et al., 2023; Billson et al., 2023). It is important to note that the WaterFormer variants were evaluated under identical experimental conditions, including dataset split, preprocessing, and training configuration. In contrast, the results of the compared methods are taken from their original publications and may not follow exactly the same experimental settings. Therefore, these

comparisons should be interpreted as indicative benchmarking rather than strictly controlled experimental equivalence.

Table 4. Results assessment of water body extraction using different remote sensing criteria

Model	Precision	Recall	F1 Score	OA	IoU	PA	UA
WaterFormer Baseline	91,2	93,2	92,2	96,8	88,7	93,2	91,2
WaterFormer -DS	92,3	93,3	92,8	97	89	93,3	92,3
WaterFormer -MS	93,2	94,2	93,7	97,4	89,7	94,2	93,2
Jagruth 2021	69	74	/	76	/	/	/
Parajuli 2023	/	83	/	89	/	/	/
George 2023	/	/	47	/	/	/	/
Billson 2023	73	68	/	/	75	/	/

Under these conditions, WaterFormer-MS shows consistently higher performance in terms of accuracy and recall, particularly in distinguishing water from spectrally similar non-water regions, although the comparison with external methods should be interpreted with caution due to differences in evaluation protocols and incomplete reporting of metrics in some referenced studies.

As shown in Figure 11, our method consistently achieves higher performance across multiple evaluation metrics, highlighting its effectiveness for robust and precise waterbody extraction.

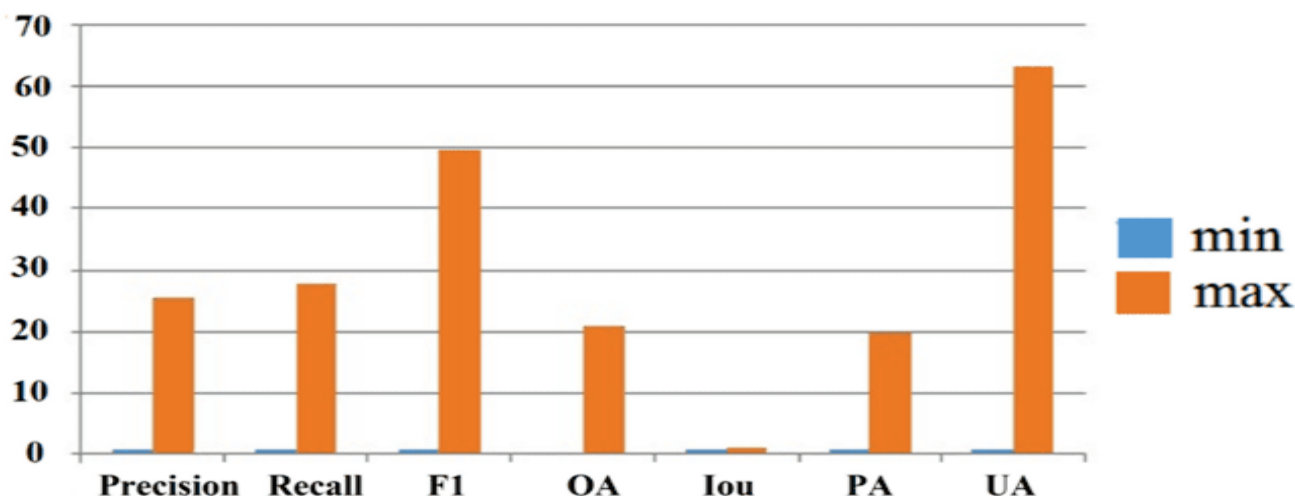


Figure 11. Results improvement with various evaluation criteria for Multi-stream method compared to other approaches

Figure 11 quantifies the relative performance of WaterFormer-MS compared to other water extraction methods. Minimum values represent the difference from the best competing method, and maximum values indicate the differences from the worst performing method. In terms of accuracy, WaterFormer-MS improves from 0.97 to 25.56, while recall improves from 0.96 to 27.79. The improvement in the F1 score ranges from 0.96 to 49.84, demonstrating its superiority over the best-performing reference methods and significant gains compared to the worst-performing methods. Overall accuracy ranges from 0.41 to 20.98, and the IoU ranges from 0.78 to 1.11, reflecting better spatial fit. The improvements in producer accuracy (PA) and user accuracy (UA) range from 0.96 to 19.94 and from 0.97 to 63.52, respectively. These results confirm that WaterFormer-MS consistently outperforms existing methods, highlighting its robustness and efficiency in accurately extracting water mass.

3.1. Technical discussion

WaterFormer is a progressive, three-variant approach for surface water extraction from satellite imagery, enabling controlled analysis of performance gains versus architectural complexity.

- V1: Lightweight encoder-decoder focused on spectral features; effective in homogeneous aquatic areas, but limited in complex environments.
- V2 (Dual Stream): Separates spectral and spatial feature extraction, preserving fine water boundaries while modeling a broader context; improves robustness in narrow and fragmented aquatic structures.
- V3 (Multi-Scale Dual Stream): Adds multi-scale feature aggregation to capture spatial context at multiple resolutions, enabling superior delineation in heterogeneous landscapes.

WaterFormer was trained on multispectral images featuring mixed urban-water interfaces, seasonal variations, and turbid or shallow waters. The multi-scale V3 version reduces false positives in complex environments while maintaining high sensitivity in narrow and discontinuous aquatic structures.

The evaluation indicators include accuracy, recall, F1 score, OA, IoU, PA, and UA. The results show a clear hierarchy from V1 to V3, corroborated by qualitative analyses of false positives and false negatives.

Although WaterFormer-MS achieved the best performance among the evaluated variants on the datasets considered in this study, the reported improvements should be interpreted within the scope of the

experimental setup. The quantitative results consistently demonstrate that combining spectral and geometric feature extraction improves segmentation accuracy across multiple evaluation metrics, including Accuracy, F1-score, IoU, PA, and UA. Furthermore, the qualitative analysis reveals a reduction in false positives and false negatives compared with the baseline models. Nevertheless, the comparisons are limited to the datasets and benchmark methods included in this study. Additional evaluations on larger and more diverse geographic regions, seasonal conditions, and independent datasets would be beneficial to further assess the generalization capability of the proposed framework. Therefore, the present results demonstrate the effectiveness of WaterFormer-MS under the tested conditions while providing a basis for future large-scale validation.

Regarding model complexity, the dual-stream architecture (V2) contributes the most significant performance improvement by explicitly separating spectral and geometric feature learning. The additional contextual and multi-scale components in V3 provide further but more incremental gains, particularly in boundary refinement and small-object preservation. This indicates that most of the performance gain originates from feature disentanglement rather than increased depth alone.

From a computational perspective, the transition from V1 to V3 introduces additional parameters and higher computational cost due to multi-stream processing and attention modules. However, the observed gains in segmentation accuracy suggest that this increase in complexity is justified in challenging heterogeneous environments, although it may not be necessary for simpler scenes.

Finally, a qualitative analysis under reduced-resolution inputs indicates that all variants experience a performance drop, especially in

boundary precision. Nevertheless, V3 demonstrates higher robustness under low-resolution conditions, suggesting that multi-scale and contextual modeling help mitigate spatial information loss.

3.2. Technical Limitations and Improvement Perspectives

The current WaterFormer method uses only optical imaging; integrating SAR data could further improve its performance in cloudy or low-contrast environments, where optical sensors may be partially affected by atmospheric conditions. While the proposed model demonstrates robustness under moderate environmental variability, its performance under severe cloud coverage conditions has not been explicitly evaluated in this study.

Incorporating attention mechanisms and adaptive normalization of multispectral data could further improve feature discrimination and generalization. Future work will focus on multimodal fusion, real-time inference, and large-scale cloud deployment for hydrological monitoring, flood mapping, and climate-resilient water resource management. While these directions increase the computational complexity of the framework, they remain feasible for offline and cloud-based processing scenarios. However, real-time inference is still a challenging objective and would require additional model optimization strategies such as pruning, quantization, or lightweight architectural redesign.

Although the proposed framework shows promising adaptability, its scalability across different geographic regions and environmental conditions has not been exhaustively validated. Therefore, further

experiments on diverse datasets from multiple regions are necessary to fully confirm its generalization capability in real-world applications.

WaterFormer's modular design promotes potential scalability and adaptability to next-generation water extraction systems, although this should be interpreted as a design characteristic rather than a fully validated property.

4. Conclusions

In this study, we introduced WaterFormer, a progressive multi-stream deep learning method for the accurate and robust extraction of water bodies from high-resolution satellite imagery. This framework comprises three complementary variants: WaterFormer-V1 (single-stream baseline model), WaterFormer-V2 (dual-stream spectral-geometric model), and WaterFormer-V3 (multi-stream model with attention and contextual modeling), enabling a systematic evaluation of the trade-offs among model complexity, segmentation accuracy, and operational feasibility.

Experimental results show that WaterFormer consistently outperforms traditional spectral index-based methods and existing deep learning approaches. Version V3 (WaterFormer-MS) achieved 93.2% precision, 94.2% recall, 93.7% F1 score, 97.4% overall accuracy, 89.7% IoU, 94.2% PA, and 93.2% UA, significantly outperforming versions V1 and V2 as well as traditional methods. Differences from previous approaches range from 0.41% to 63.5%, highlighting the robustness of the proposed multi-stream and multi-scale design under the evaluated experimental conditions. Version V2 offers balanced improvements in boundary preservation and contextual modeling, while version V1 serves

as a lightweight baseline suitable for low-computational-resource scenarios.

Qualitative analyses confirm that WaterFormer produces spatially consistent water masks with fewer false positives and false negatives, even in complex environments such as urban-water interfaces, narrow streams, fragmented surfaces, and areas of seasonal variation. By jointly leveraging multiscale spectral, spatial, and contextual information, the approach demonstrates consistent performance across the tested datasets and regions, while further validation on additional and more diverse datasets is still required to fully establish its generalization capability.

From an application perspective, the proposed framework has potential for hydrological monitoring, flood risk assessment, and climate-responsive water resource management, including reservoir monitoring, irrigation management, and agricultural water-use analysis. However, these applications remain prospective and require dedicated validation studies before operational deployment.

4.1. Future research directions include

- Multimodal fusion of optical and SAR data to improve performance under cloud-covered conditions
- Incorporation of advanced attention mechanisms to further enhance feature representation
- Temporal modeling for multi-date water dynamics analysis
- Optimization for deployment on large-scale or resource-constrained environments

In summary, WaterFormer-MS provides an effective multi-stream architecture for surface water segmentation, with demonstrated improvements in the evaluated experimental setting. Further research is required to assess its generalization capability across additional datasets, geographic regions, environmental conditions, and broader operational monitoring scenarios.

5. Acknowledgments

The authors would like to thank their institutions and colleagues for their support during the development of this work. The authors also acknowledge the publicly available datasets and resources that contributed to the experimental study.

6. Funding

This research received no specific external funding.

7. Statement of conflict of interest

The authors declare that they have no conflict of interest related to this work.

8. Statement on the use of IA

ChatGPT (OpenAI) was used only to assist with language editing, grammar checking, and improving the clarity of the manuscript. No AI-based or automated tools were used to manage the references. The references were managed and verified manually by the authors. The

authors reviewed and verified all AI-assisted content and remain fully responsible for the accuracy, originality, and final version of the manuscript.

9. References

- Ambrosio-Ambrosio, J. P., & González-Camacho, J. M. (2023). Algoritmos de aprendizaje automático para clasificar zonas de inundación a partir de imágenes de radar de apertura sintética. *Tecnología y ciencias del agua*, 14(4), 107–154. <https://doi.org/10.24850/j-tyca-14-04-03>
- Anandakrishnan, J., Sundaram, V. M., & Paneer, P. (2024). CERMNet: A SAR-optical feature fusion for cloud elimination from Sentinel-2 imagery using residual multiscale dilated network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17, 11741–11749. <https://doi.org/10.1109/JSTARS.2024.3411032>
- Anandakrishnan, J., Sundaram, V. M., & Paneer, P. (2025). STA-AgriNet: A spatio-temporal attention framework for crop type mapping from fused multi-sensor multi-temporal SITS. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18, 1817–1826. <https://doi.org/10.1109/JSTARS.2024.3510468>
- 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. <https://doi.org/10.1109/TPAMI.2016.2644615>

- Benali, A. (2024). Water bodies extraction using mathematical morphology. *Space Science and Technology*, 30(4), 48–57. <https://doi.org/10.15407/knit2024.04.048>
- Benali, A., Kharbouch, H., Krachai, M. D., Singh, J. P., Bouddou, R., Bendaha, Y., Belabbes, A., Husein, A., Salau, A. O. & Rubanenko, O. (2025). Dual-Stream convolutional networks for scalable and precise water body mapping from multispectral Earth observation imagery. *International Journal of Computational Intelligence Systems*, 18, 260. <https://doi.org/10.1007/s44196-025-00985-3>
- Billson, J., Islam, M. S., Sun, X., & Cheng, I. (2023). Water body extraction from Sentinel-2 imagery with deep convolutional networks and pixelwise category transplantation. *Remote Sensing*, 15(5), 1253. <https://doi.org/10.3390/rs15051253>
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Computer vision – ECCV 2018* (pp. 833–851). Springer. https://doi.org/10.1007/978-3-030-01234-2_49
- Chen, Y., Tang, L., Kan, Z., Bilal, M., & Li, Q. (2020). A novel water body extraction neural network (WBE-NN) for optical high-resolution multispectral imagery. *Journal of Hydrology*, 588, 125092. <https://doi.org/10.1016/j.jhydrol.2020.125092>
- Flores-Escobar, F., Palacios-Vélez, E., Exebio-García, A. A., Fernández-Reynoso, D. S., & Valdivia-Alcalá, R. (2025). Estimación de volumen de una presa usando percepción remota y Machine Learning. *Tecnología y ciencias del agua*, 16(6), 43–79. <https://doi.org/10.24850/j-tyca-2025-06-02>

- George, B., Sajith, V. V., Sowmya, V., & Sivanpillai, R. (2023). Performance improvement of water body segmentation by DeepLabV3+ using two-dimensional variational mode decomposition. In *2023 10th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 603–608). IEEE. <https://doi.org/10.1109/SPIN57001.2023.10116311>
- Guo, H., He, G., Jiang, W., Yin, R., Yan, L., & Leng, W. (2020). A multi-scale water extraction convolutional neural network (MWEN) method for GaoFen-1 remote sensing images. *ISPRS International Journal of Geo-Information*, 9(4), Article 189. <https://doi.org/10.3390/ijgi9040189>
- Jing, B., & Ning, W. (2021). 基于阈值分割与决策树的SAR影像水体信息提取 [Water body information extraction of SAR images based on threshold segmentation and decision tree]. *Geospatial Information*, 19(3), 46–49. <https://dxkj.cbpt.cnki.net/portal/journal/portal/client/paper/6b842eaccd0767ec9dfe6416af35d2eb>
- Jagruth, K., Manikandan, V. M., & Kumar, R. K. (2021). Water body identification from the satellite images using color component analysis with morphological operations. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ICCCNT51525.2021.9579839>

- Wang, F., Zhou, Q., Gao, H., Wen, Y., & Zhou, S. (2024). Dynamic monitoring of Poyang Lake water area and storage changes from 2002 to 2022 via remote sensing and satellite gravimetry techniques. *Remote Sensing*, 16(13), 2408. <https://doi.org/10.3390/rs16132408>
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- Mondejar, J. P., & Tongco, A. F. (2019). Near infrared band of Landsat 8 as water index: A case study around Cordova and Lapu-Lapu City, Cebu, Philippines. *Sustainable Environment Research*, 29(1), 16. <https://doi.org/10.1186/s42834-019-0016-5>
- Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M. C. H., Heinrich, M. P., Misawa, K., Mori, K., McDonagh, S. G., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv*. <https://doi.org/10.48550/arXiv.1804.03999>
- Parajuli, J., Fernandez-Beltran, R., Kang, J., & Pla, F. (2022). Attentional dense convolutional neural network for water body extraction from Sentinel-2 images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 6804–6816. <https://doi.org/10.1109/JSTARS.2022.3198497>
- Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>

- Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>
- Tacsi-Palacios, A., Mejía, A., Condom, T., & Cardenas-Gaudry, M. (2025). La contribución hídrica del glaciar y la dinámica de los procesos glaciológicos-hidrológicos de la cuenca. Caso Yanamarey, Cordillera Blanca, Perú. *Tecnología y ciencias del agua*, 16(6), 01–42. <https://doi.org/10.24850/j-tyca-2025-06-01>
- Tian, B., Zhang, F., Lang, F., Wang, C., Wang, C., Wang, S., & Li, J. (2022). A Novel Water Index Fusing SAR and Optical Imagery (SOWI). *Remote Sensing*, 14(21), 5316. <https://doi.org/10.3390/rs14215316>.
- Wang, G. H., Pei, L., Du, Q. Y., & Li, X. (2020). 针对资源三号卫星影像水体提取的谱间关系法 [Spectral relation method for water body extraction from ZY-3 imagery]. *Remote Sensing Information*, 35(3), 117–121. <https://doi.org/10.3969/j.issn.1000-3177.2020.03.016>
- Wang, N., Cheng, J., Zhang, H., Cao, H., & Liu, J. (2020). Application of U-net model to water extraction with high resolution remote sensing data. *Remote Sensing for Natural Resources*, 32(1), 35–42. <https://www.gtzyyg.com/en/article/doi/10.6046/gtzyyg.2020.01.06>
- Wang, R., Li, D., & Liu, B. (2022). A water body identification model for lakes in Hoh Xil based on GF-6 WFV satellite data. *Remote Sensing of Natural Resources*, 34(2), 80–87. <https://doi.org/10.6046/zrzyyg.2021125>

- Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), 3025–3033. <https://doi.org/10.1080/01431160600589179>
- Yu, Y., Yao, Y., Guan, H., Li, D., Liu, Z., Wang, L., Yu, C., Xiao, S., Wang, W., & Chang, L. (2021). A self-attention capsule feature pyramid network for water body extraction from remote sensing imagery. *International Journal of Remote Sensing*, 42(5), 1801–1822. <https://doi.org/10.1080/01431161.2020.1842544>
- Zhao, F., Feng, S., Xie, F., Zhu, S., & Zhang, S. (2023). Extraction of long time series wetland information based on Google Earth Engine and random forest algorithm for a plateau lake basin—A case study of Dianchi Lake, Yunnan Province, China. *Ecological Indicators*, 146, 109813. <https://doi.org/10.1016/j.ecolind.2022.109813>
- Zhang, D., Yang, S., Wang, Y., & Zheng, W. (2019). Refined water body information extraction of Three Gorges reservoir by using GF-1 satellite data. *Yangtze River*, 50(9), 233–239. <https://doi.org/10.16232/j.cnki.1001-4179.2019.09.039>
- Ji, S., Wei, S., & Lu, M. (2019). Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 574–586. <https://doi.org/10.1109/TGRS.2018.2858817>

- Zhou, T., Wang, Y., Zou, J., Li, C., Cui, Y., Wang, X., Xie, C., & Xia, P. (2023). PCA and SVM-based algorithm of water area extraction from remote sensing images and its verification. *Water Resources Protection*, 39(2), 180–189. DOI: 10.3880/j.issn.1004-6933.2023.02.022
- Zhong, H.-F., Sun, H.-M., Han, D.-N., Li, Z.-H., & Jia, R.-S. (2022). Lake water body extraction of optical remote sensing images based on semantic segmentation. *Applied Intelligence*, 52(15), 17974–17989. <https://doi.org/10.1007/s10489-022-03345-2>
- Wang, X., Fu, X., Hu, H., & Li, H. (2023). Research on water extraction method from remote sensing images of lakes in cold and arid regions based on deep learning. In *Proceedings of the 3rd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2023)*. SPIE. <https://doi.org/10.1117/12.2685343>