A Radiometrically-Agnostic SegFormer Framework for Water Segmentation in C-Band SAR Imagery
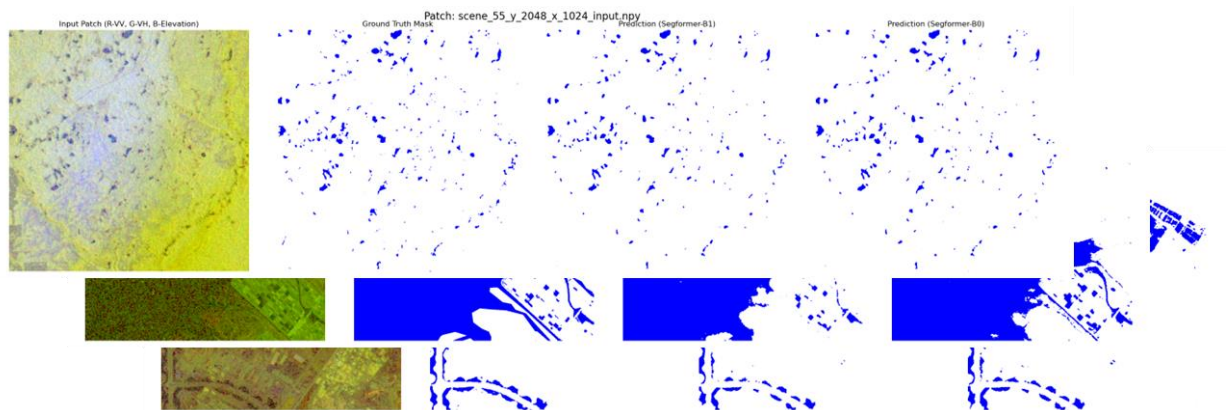
Algorithm Theoretical Basis Document (ATBD)

Document Control Sheet

1	Security Classification	Unrestricted		
2	Distribution			
3	Report / Document version	(a) Issue no. 1	(b) Revision & Date	1.0 & 10-10-2025
4	Report / Document Type	Technical Document		
5	Document Control Number	NRSC-RSA-WATER RE-WRD-OCT 2025-TD-0002915-V1.0		
6	Title	A Radiometrically-Agnostic SegFormer Framework for Water Segmentation in C-Band SAR Imagery Algorithm Theoretical Basis Document (ATBD)		

7	Particulars of collation	Pages 22	Figures 6	Tables 3	References 38

8	Author(s)	Bharath Kumar Reddy K		
9	Affiliation of authors	Water Resources Group (WRG)		
10	Project Team	-		
11	Scrutiny mechanism	Compiled by Bharath Kumar Reddy K	Reviewed by Group Head, WRG	Approved / Controlled by DD, RSAA
12	Originating unit	National Remote Sensing Centre		
13	Sponsor(s) Name	-NA-		
14	Date of Initiation	01-Sep-2025		
15	Date of Publication			

16	**Abstract (with Keywords):** Accurate and timely mapping of surface water bodies is critical for sustainable water management, yet automated methods must overcome challenges posed by diverse sensor characteristics and complex environmental conditions. This paper presents a robust and efficient end-to-end framework for water segmentation that addresses these issues. We adapt the lightweight SegFormer Mix-Vision Transformer (MiT) architecture for multi-modal data fusion, combining C-band Synthetic Aperture Radar (SAR)

imagery with topographic features (elevation and slope). A key contribution is our per-scene normalization strategy, which renders the model agnostic to the input's radiometric correction (e.g., Sigma, Beta, or Gamma Naught), significantly enhancing its generalization across different sensors. For rigorous validation, we introduce a dual-stratified evaluation protocol. Through a comparative analysis of SegFormer B0 and B1 variants, we demonstrate that the more efficient B0 model achieves state-of-the-art performance, attaining a Water IoU of 0.958 on a held-out test set, slightly surpassing the larger B1 model while being approximately 48% faster at inference. Our findings indicate that for this application, a well-adapted, lightweight transformer provides the optimal balance of performance and computational efficiency, offering a pragmatic and highly accurate solution for operational, large-scale water monitoring.

1. Introduction

Despite Earth's appearance as a 'blue planet,' the minute fraction of its water that is fresh and accessible underscores a global imperative for its effective management (Gupta et al., 2021). Central to this effort is the rapid and accurate mapping of surface water bodies, which are foundational to ecosystem stability, human welfare, and biodiversity (Kadapala et al., 2024; Vörösmarty et al., 2010). The ability to precisely delineate and monitor these resources is not merely a technical exercise but a prerequisite for sustainable land and water management on both local and global scales (Wang et al., 2021). This need is formally recognized by the United Nations' Sustainable Development Goal (SDG) 6, which targets sustainable water management by 2030 (Druce et al., 2021; Hakimdavar et al., 2020). Therefore, the development of robust methodologies that can capture the spatial and temporal dynamics of surface water is a foundational step toward addressing the global water crisis and achieving sustainability goals (Nagaraj et al., 2024).

To transcend the spatial and temporal limitations of traditional monitoring techniques (Palmer et al., 2015), the scientific community has widely adopted remote sensing for a synoptic view of Earth's water bodies (Meng et al., 2012; Woolway et al., 2020; Deke et al., 2025). This technology provides an unparalleled vantage point, proving especially effective for mapping waterbodies at scale where ground-based campaigns are unfeasible (Kalpan et al., 2018). For decades, optical sensors, such as those in the Landsat series of satellites and Sentinel-2A/B satellites recently, have been the workhorse for these applications, leading to numerous successful mapping studies (Dan et al., 2020; Pekel et al., 2016; Kadapala et al. 2023). Their popularity is driven by high spatial resolutions and rich spectral information that can detail water extent on a large scale. Despite these significant advantages, optical sensors possess critical operational constraints: their reliance on passive sunlight renders them ineffective in the presence of cloud cover or at night.

The challenge of cloud cover and nighttime limitations in optical remote sensing has driven significant research into Synthetic Aperture Radar (SAR) for all-weather, day-and-night water body monitoring. Early approaches, often termed "traditional methods," relied on manually engineered features and statistical properties. These included edge detection algorithms like Canny and Sobel (Chen et al., 2021)[16, 21], variational methods such as Active Contour Models (ACMs) and Level Sets (Silviera et al., 2008; Meng et al., 2019), region growing (Kadapala et al., 2024; Liu et al., 2016)clustering techniques like K-Means and Fuzzy C-Means (Zheng et al., 2014; Li Ning et al., 2019), and early machine learning classifiers like Support Vector Machines (SVMs) and Random Forests (RFs) (Lv et al., 2010; Cafero et al., 2012). Edge detection techniques are simple but prone to discontinuities and noise. Level set and active contour models offer flexible and accurate contour detection but are computationally expensive and require careful

initialization. Machine learning approaches like clustering, Random Forests, and SVMs provide robustness and efficiency but can suffer from high computational costs and sensitivity to outliers. Meanwhile, Markov Random Field and statistical algorithms effectively incorporate spatial and contextual information but are slow and highly dependent on the underlying statistical models and handcrafted features (Guo et al., 2022).

The advent of deep learning, particularly Convolutional Neural Networks (CNNs), marked a paradigm shift, moving from handcrafted features to automated feature learning. The progression evolved from early patch-based CNN classifiers to modern end-to-end, fully convolutional architectures (Long et al., 2015; Ren et al 2020). The U-Net architecture (Ronneberger et al., 2015), with its encoder-decoder structure and characteristic skip connections, has become the dominant paradigm in this domain. Its ability to fuse high-resolution spatial details from shallow layers with deep semantic features makes it exceptionally well-suited for preserving the precise boundaries of water bodies, leading to numerous successful applications and modifications (Katiya et al., 2021; Konapala et al., 2021; Kim et al., 2021).

Subsequent research has largely focused on enhancing the U-Net framework. These advancements include the integration of attention mechanisms to help the model focus on salient features (e.g., DAU-Net, LGFUNet [Ren et al., 2021; Bai et al., 2025]), the use of multi-scale feature extraction modules like Atrous Spatial Pyramid Pooling (ASPP) to broaden the model's receptive field (Chen et al., 2018), and the fusion of SAR data with other sources like optical imagery (Konapala et al. 2021). However, despite these improvements, CNN-based architectures are fundamentally limited by the local nature of their convolutional kernels. They often struggle to capture the long-range spatial dependencies and global context across a large, complex scene, which is critical for distinguishing between spectrally similar features (Ding et al., 2022; Wang et al., 2022). Furthermore, many existing deep learning approaches are sensitive to the absolute backscatter values of the input data, requiring consistent radiometric correction and calibration (Arora et al., 2025).

To address these limitations, our work moves beyond CNNs to leverage the power of Vision Transformers. We propose a robust, efficient, and generalizable end-to-end framework built upon a lightweight SegFormer architecture. Segfomers are relatively new are being increasingly used in various earth observation applications (Lin et al., 2023; Li et al., 2023). The primary novelty of this work lies in its radiometric robustness, achieved through a per-scene normalization strategy that makes the model agnostic to the specific radiometric calibration of input SAR data (e.g., Sigma Naught, Beta Naught, or Gamma Naught) and the sensor used for data acquisition. Building on this foundation, we introduce the framework that extends a pre-trained transformer to jointly process C-band

SAR and topographic datasets (elevation and slope), thereby enhancing its ability to generalize and operate effectively across diverse real-world scenarios.

This work also addresses two critical aspects of operational deployment: robust evaluation and inference. We introduce a rigorous dual-stratified evaluation protocol to provide both a balanced assessment across all water-cover conditions and a realistic estimate of real-world performance. For producing seamless, large-scale maps, we implement and systematically compare single-pass (non-overlapping) and multi-pass (overlapping) tiled inference strategies, quantifying the impact of multi-context probability averaging on the final segmentation accuracy.

2. Materials and Methods

2.1. Datasets

To develop a robust and generalizable model, this study utilizes a combination of a large, publicly available global dataset for primary training and a distinct in-house dataset for supplementary validation.

The primary dataset is the S1S2-Water global dataset, a publicly available resource designed for semantic segmentation of water bodies (Wieland et al., 2023). This dataset consists of 65 globally distributed scenes, each providing a 10-meter resolution Sentinel-1 C-band Synthetic Aperture Radar (SAR) image, a corresponding ground truth binary water mask, and a Digital Elevation Model (DEM) from the Copernicus DEM GLO-30. For this study, we utilized the dual-polarization (VV/VH) Sentinel-1 SAR data, the DEMs, and the binary water masks to form our multi-modal input.

To improve the model's generalization capabilities on a different sensor, we also incorporated an in-house dataset acquired by India's EOS-04 (RISAT-1A) satellite. This dataset consists of C-band SAR imagery from the Multi-Resolution Scanner (MRS) at an 18-meter spatial resolution, with different sensor characteristics, noise profiles, and spatial resolution enhancing the generalization   capability of the model. Both Sigma Naught and Gamma Naught images from EOS-04 are used in the model training. The EOS-04 datasets were treated with standard pre-processing techniques (Kadapala et al., 2024)

2.2. Data Pre-processing and Patch Generation

A robust, multi-stage pre-processing pipeline was developed to convert the raw, variable-sized satellite scenes into a uniform dataset suitable for deep learning. To accommodate the different characteristics of the public and in-house datasets, two distinct processing routines were implemented.

For the in-house EOS-04 dataset, which was paired with large-area Copernicus DEM, a more intensive on-the-fly workflow was required. For each 18-meter resolution EOS-04 scene, the corresponding elevation and slope layers were generated. Then, both the DEM and its derived slope raster were co-registered to the specific grid of the EOS-04 SAR image using GDAL's Warp function with Bilinear interpolation. Any NaN values present in the raw SAR bands were imputed with a low backscatter value of -25.0, consistent with water-like signatures. The two bands of the SAR image were then stacked with the co-registered elevation and slope layers to form a 4-channel input array for that scene.

For the S1S2-Water dataset, a more direct approach was possible as the 10-meter resolution Sentinel-1 imagery and the Copernicus 30m DEM and slope were already provided as co-registered scenes. The two bands of the Sentinel-1 image (VV and VH) were then directly stacked with the corresponding elevation and slope layers to form the 4-channel input array, without the need for the more complex on-the-fly warping performed on the EOS-04 data.

Following the creation of the 4-channel scene arrays for both datasets, a common set of procedures was applied. A per-scene min-max normalization strategy was used for each of the four input channels, scaling the pixel values to a range of [0.0, 1.0] based on the minimum and maximum values of the valid pixels within that scene. The full-resolution, normalized scenes were then tiled into patches. To minimize information loss, an edge-aligned patching strategy was employed to extract 1024x1024 pixel patches. Finally, to prevent the model from learning incorrect patterns from no-data areas (defined as pixels with a 0 value in the original SAR data), a distinct out-of-band value of -1.0 was used to fill all invalid pixels within each final input patch. Corresponding pixels in the ground truth masks were assigned an ignore index of 255.

2.3. Model Architecture and Adaptation

The core of our framework is the SegFormer architecture (Figure-1), chosen for its excellent balance of performance and computational efficiency in semantic segmentation. To investigate the trade-off between model capacity and inference speed for our specific task, we selected two variants of the underlying Mix Transformer (MiT) backbone for a comparative analysis: the lightweight MiT-B0 and the larger MiT-B1. The primary architectural difference lies in the feature dimensionality; the MiT-B0 backbone uses channel dimensions [C1, C2, C3, C4] (Figure-1) of [32, 64, 160, 256] across its four encoder stages, while the larger MiT-B1 doubles these dimensions to [64, 128, 320, 512], respectively.

Both models utilize checkpoints that were first pre-trained on ImageNet and subsequently fine-tuned on the ADE20K semantic segmentation dataset. Since the models expect a 3-channel (RGB) input, and their final layer is configured for 150 classes. We adapted both architectures to our task in two key steps. First, to accommodate our

4-channel fusion of SAR and topographic data, we replaced the initial 7x7 convolutional layer in the first patch embedding stage (encoder.patch_embeddings[0].proj). The original layer with in_channels=3 was substituted with a new layer where in_channels=4. To leverage transfer learning, the weights for the first three channels were copied from the original pre-trained weights, while the weights for the new fourth channel were initialized as the mean of the original three.

The final classification layer (decode_head.classifier), originally designed for 150 classes, was automatically replaced upon loading with a new, randomly initialized 1x1 convolutional layer with out_channels=2, corresponding to our 'background' and 'water' classes. This transfer learning approach allows both models to utilize the powerful, pre-trained features from the deep transformer blocks while adapting the input and output layers to our novel application. By training and evaluating both the adapted SegFormer-B0 and SegFormer-B1 models, this study provides a direct comparison of their performance, offering insights into the optimal architectural choice for operational, large-scale water mapping.
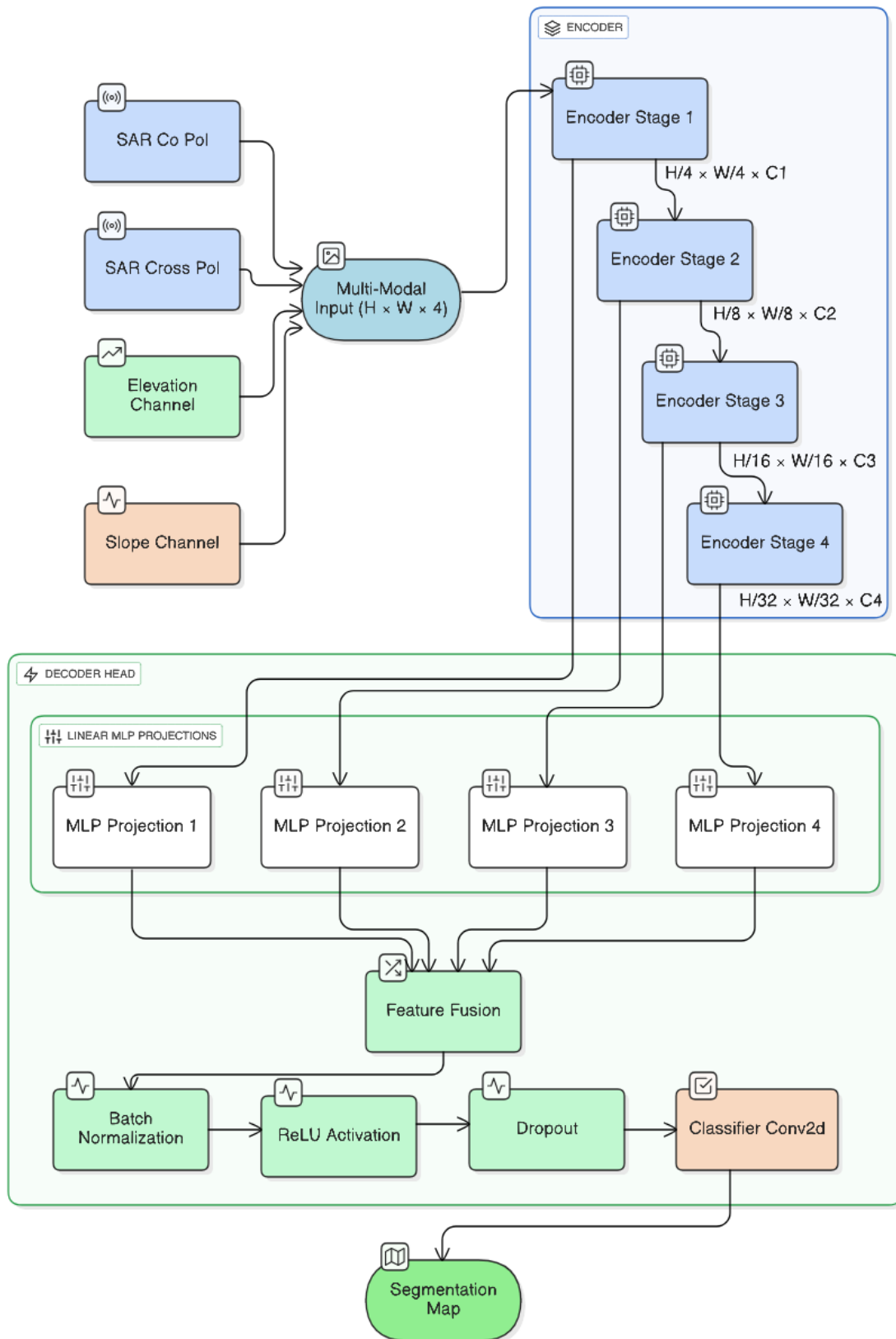
Figure 1 Segformer architecture

2.4. Training and Evaluation Protocol

The model was trained for 50 epochs using the AdamW optimizer with a learning rate of 6×10^{-5}. To accommodate the large 1024x1024 patch size on a 16GB GPU, a per_device_train_batch_size of 4 was used. Training was accelerated using automatic mixed-precision (fp16).

For evaluation, a stratified sampling methodology was employed to create a fixed-size, representative test set. All generated patches were analyzed to determine their water-cover percentage and grouped into five bins (0-20%, 20-40%, etc.). A fixed random set of 100 patches, comprising 20 randomly selected samples from each bin, was compiled to form the evaluation set, while all remaining patches were used for training. Model performance was evaluated at the end of each epoch. The primary metric for selecting the best model checkpoint was the IoU of the water class (Water IoU), supplemented by Mean IoU (mIoU) and overall pixel accuracy. The Cross-Entropy loss function was configured to ignore the index 255 during training. The overview of the entire approach is presented in Figure-2.
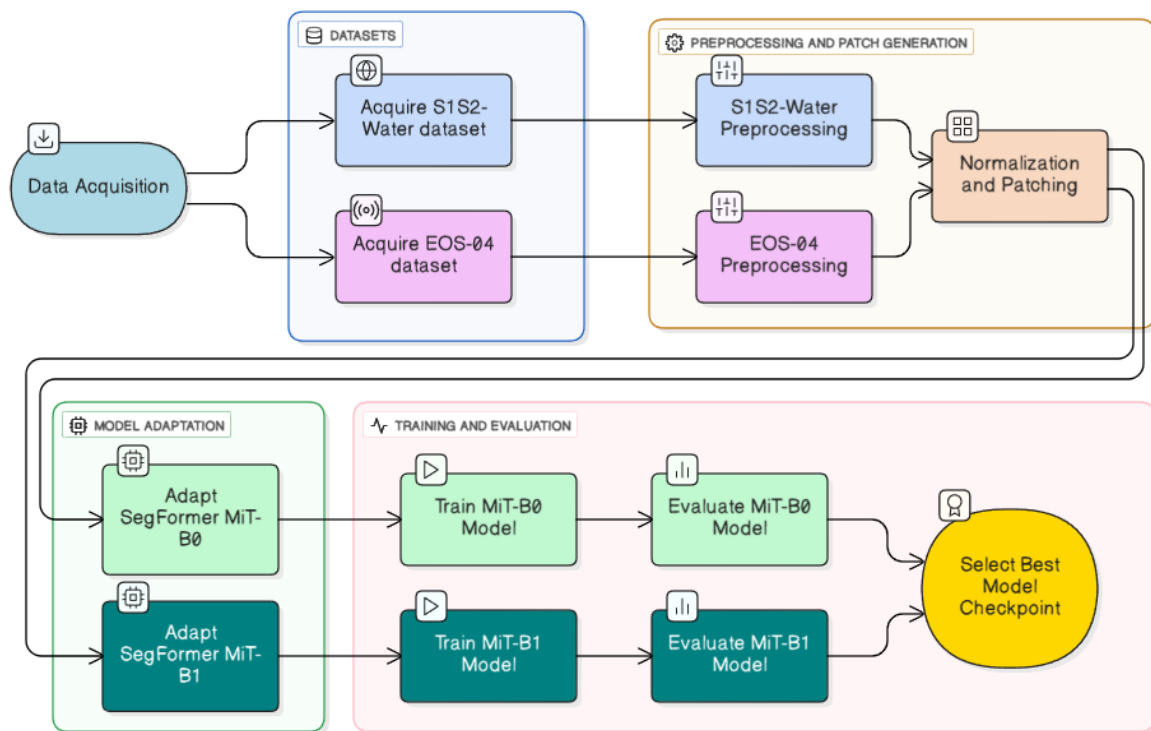

Figure 2 Overview of the approach

3. Results and Discussions

3.1. Model Training and Convergence

Both the adapted SegFormer-B0 and SegFormer-B1 models were successfully trained for 50 epochs on the multi-modal patch dataset. The training process for both models showed stable convergence, marked by a consistent decrease in training loss, which confirms the effectiveness of the transfer learning approach. The performance on the stratified validation set was evaluated at the end of each epoch, with the Water IoU (eval_water_iou) used as the primary metric for selecting the best model checkpoint.

The peak performance achieved by each model on the validation set is summarized in Table 1.

Table 1 Peak validation performance of the trained SegFormer models.

Model	Best eval_water_iou	Best Epoch	eval_accuracy (at best epoch)
SegFormer-B0	0.9582	50	0.9592
SegFormer-B1	0.9560	16	0.9581

The lightweight SegFormer-B0 model demonstrated a remarkably stable and robust learning trajectory. Its validation Water IoU steadily increased throughout the 50 epochs, achieving its peak score of 0.9582 in the final epoch. In contrast, the larger SegFormer-B1 model, with its greater capacity, exhibited faster initial convergence but also showed clear signs of overfitting in later stages. It achieved its peak validation Water IoU of 0.9560 relatively early, at epoch 16. After this point, while its training loss continued to decrease, the validation performance on unseen patches fluctuated and did not surpass this peak. This suggests that the model had started to memorize the training set at the expense of its generalization capability.

3.2. Performance on the Held-Out Dataset

To provide an unbiased assessment of the generalized performance of both trained models, we evaluated them on a held-out test patches. This consisted of 100 patches, curated via stratified random sampling to ensure a representative mix of water coverages, which were not used during the training or validation phases. We asessed both models on key segmentation metrics, including Water IoU, Producer's and User's Accuracy for the water class, and average inference speed.

The quantitative results, summarized in Table 2, confirm the findings observed during the validation phase. Both models showed similar performance across all primary accuracy metrics, while B0 being significantly more computationally efficient.

Table 2 Performance comparison on the held-out test set.

Metric	SegFormer-B0	SegFormer-B1
Water IoU	0.9580	0.9521
Mean IoU	0.9719	0.9753
Producer's Accuracy (Recall)	0.9742	0.9698
User's Accuracy (Precision)	0.9829	0.9812
Overall Accuracy	0.9937	0.9928
Avg. Inference Time (s/patch)	0.0237	0.0350

The SegFormer-B0 model achieved a Water IoU of 0.9580, compared to 0.9521 for the B1 model. Furthermore, the B0 model demonstrated a superior balance of Producer's Accuracy (Recall) and User's Accuracy (Precision) for the water class, at 0.9742 and 0.9829 respectively. This indicates a high reliability in both identifying the vast majority of true water pixels on the ground (high recall) and ensuring that the pixels it classifies as water are indeed water (high precision).

In terms of computational efficiency, the advantages of the smaller B0 architecture are substantial. The SegFormer-B0 model processed each 1024x1024 patch in an average of 0.0237 seconds, while the B1 model required 0.0350 seconds. This makes the B0 model approximately 48% faster at inference, a critical advantage for processing large satellite scenes in an operational context.

## 3.3. Impact of Multi-Pass Inference on EOS-04 Beta Naught image

To assess whether prediction averaging could mitigate classification uncertainties at tile boundaries and to test the model's performance on a sensor not seen during training, we evaluated a multi-pass, overlapping inference strategy. This validation was performed exclusively on a held-out test set derived from an RISAT-1A (EOS-04) scene. The input C-band SAR data for this experiment was a Beta Naught backscatter product, which had subjected for noise removal and scaled to decibels (dB). We compared the performance of the multi-pass method against a standard single-pass (non-overlapping) approach for both the B0 and B1 models on this dataset. The classified water layer is shown in Figure 3 and the results are presented in Table 3.
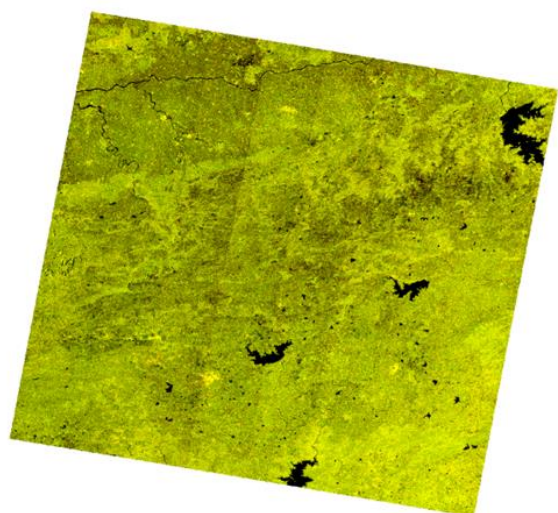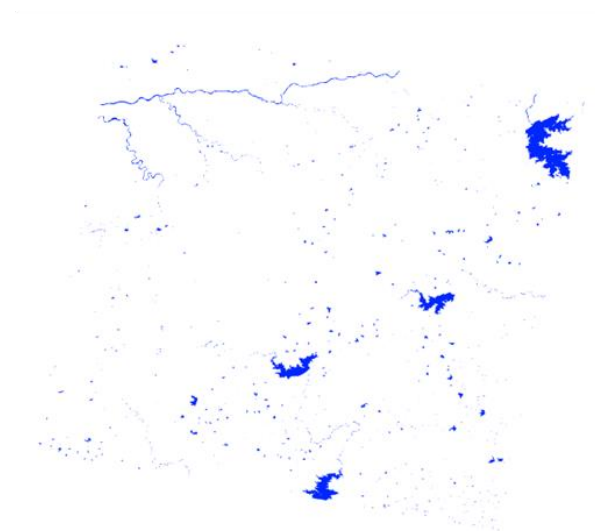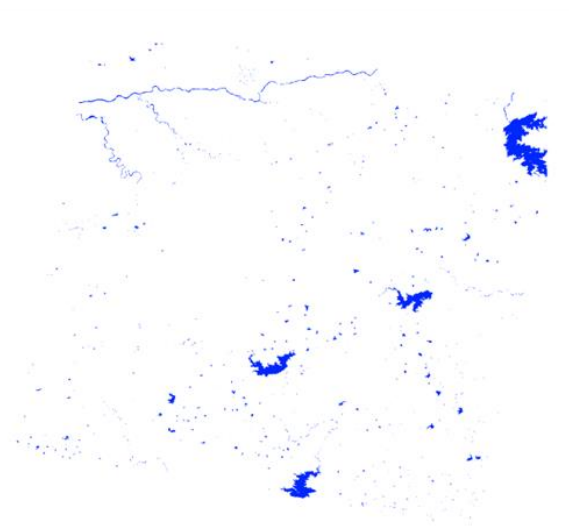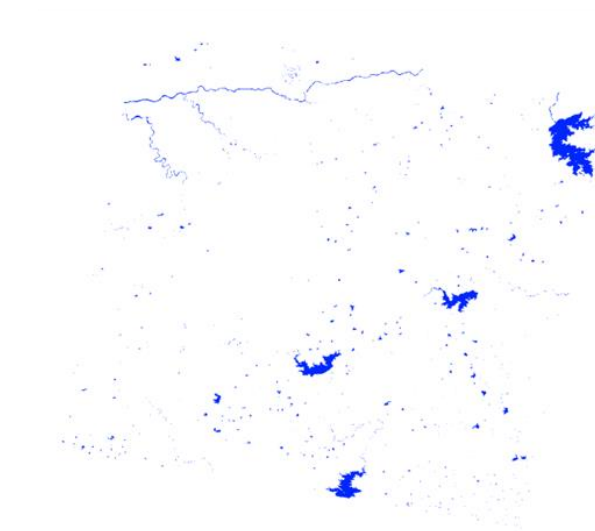
(a) SAR Image
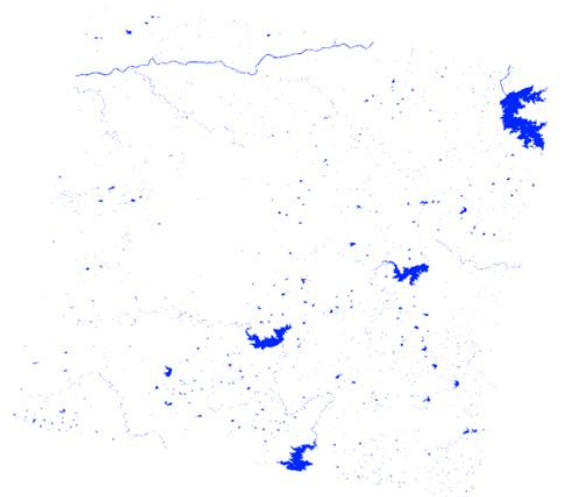
(b) Water reference mask
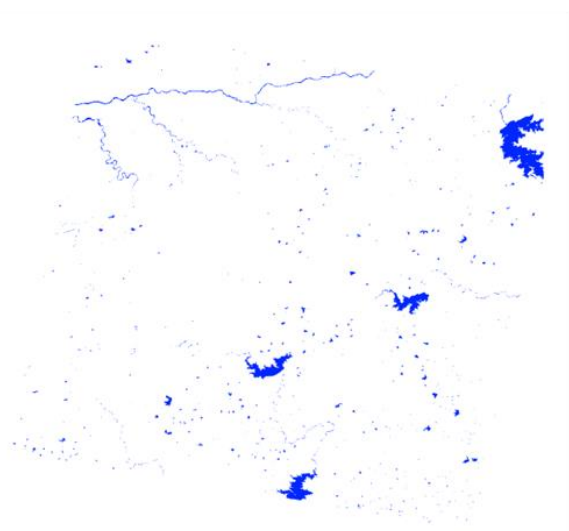
(c) Water - B0 architecture – 1 inference

(d) Water - B0 architecture – 3 inferences

(e) Water – B1 architecture – 1 inference

(f) Water – B1 architecture – 3 inferences

Figure 3 Water layer extracted using B0 and B1 architectures with single and multiple inferences

The analysis indicates that the application of multi-pass inference provides a slight, but not statistically significant, improvement in segmentation accuracy on the EOS-04 data. For the SegFormer-B0 model, the Water IoU increased marginally from 0.8255 with a single pass to 0.8307 with three overlapping passes. While this confirms that allowing the model to see a region in multiple contexts can help resolve some ambiguities, the modest nature of the gain suggests that for this application, a single-pass inference strategy is largely sufficient and offers a faster processing time.

Table 3 Performance of models on a beta naught product with single and multi-inference strategy

Model	Inference Passes	Water IoU	Mean IoU	Producer's Accuracy	User's Accuracy	Overall Accuracy
SegFormer-B0	1	0.8255	0.9109	0.8938	0.9153	0.9963
	3	**0.8307**	**0.9135**	**0.8975**	**0.9177**	**0.9964**
SegFormer-B1	1	0.8213	0.9087	0.8667	0.9400	0.9963
	3	**0.8231**	**0.9097**	**0.8683**	**0.9405**	**0.9963**

3.4 Strengths and Limitations

A qualitative assessment of the classification results was conducted to understand the operational strengths and weaknesses of the proposed algorithm. The model demonstrated remarkable generalization across data from multiple C-band SAR sensors (Sentinel-1 and EOS-04), despite their varying noise characteristics and spatial resolutions. This robustness is attributed to the per-scene normalization strategy, which renders the model's performance largely agnostic to the input backscatter coefficient (i.e., Sigma Naught, Beta Naught, or Gamma Naught). The framework proved highly effective in challenging scenarios where traditional algorithms often fail. For instance, the model successfully delineated small (Figure 5), intricate water bodies, such as narrow river networks. Furthermore, it demonstrated robustness to confounding surface features, successfully classifying water bodies even when covered by a thin ice sheet (Figure 4). Despite the ice altering the C-band backscatter signature to be distinct from open water, the model was able to leverage the combined context from the topographic channels and surrounding spatial patterns to correctly identify the underlying class as water. This ability to overcome backscatter ambiguity is a significant challenge for traditional algorithms. This success is likely due to the SegFormer architecture's ability to leverage global spatial context from its transformer backbone, allowing it to differentiate features that are spectrally similar but contextually distinct.
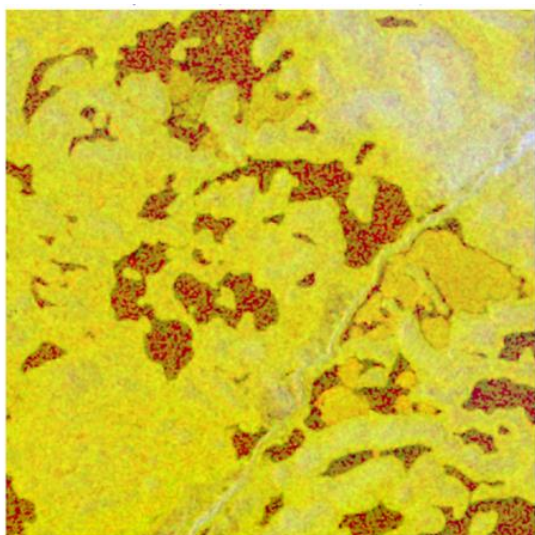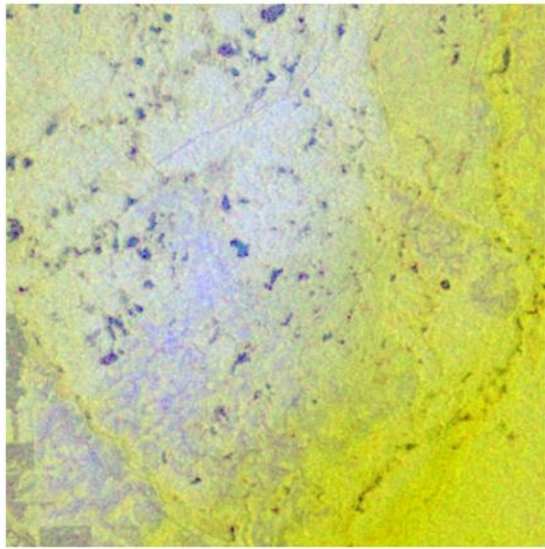
(a) SAR Image

(b) Water reference mask
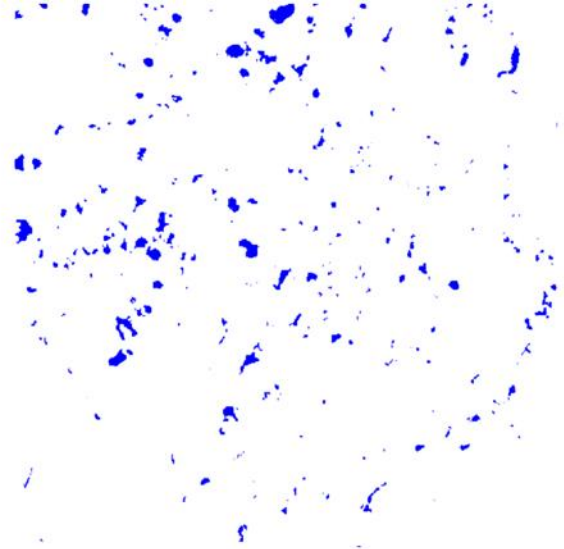
(c) Water extraction using B1 architecture

(d) Water extraction using B0 architecture

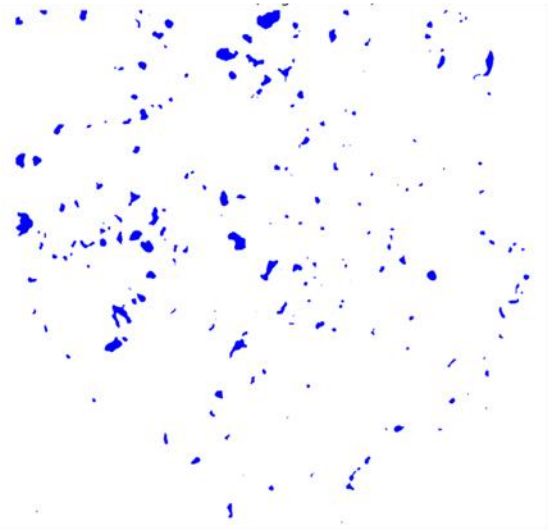Figure 4 Performance of algorithm on water covered by thin ice

(a) SAR Image   (b) Water reference mask

(c) Water extraction using B1 architecture   (d) Water extraction using B0 architecture

Figure 5 Performance of algorithm over small water bodies

Despite the excellent overall performance, the model exhibited some inconsistencies in complex water-land transition zones (Figure 6). These instances of uncertainty highlight a potential avenue for future work, which could involve the integration of additional data modalities or the development of a multi-class segmentation scheme to explicitly identify these transitional wetland classes.
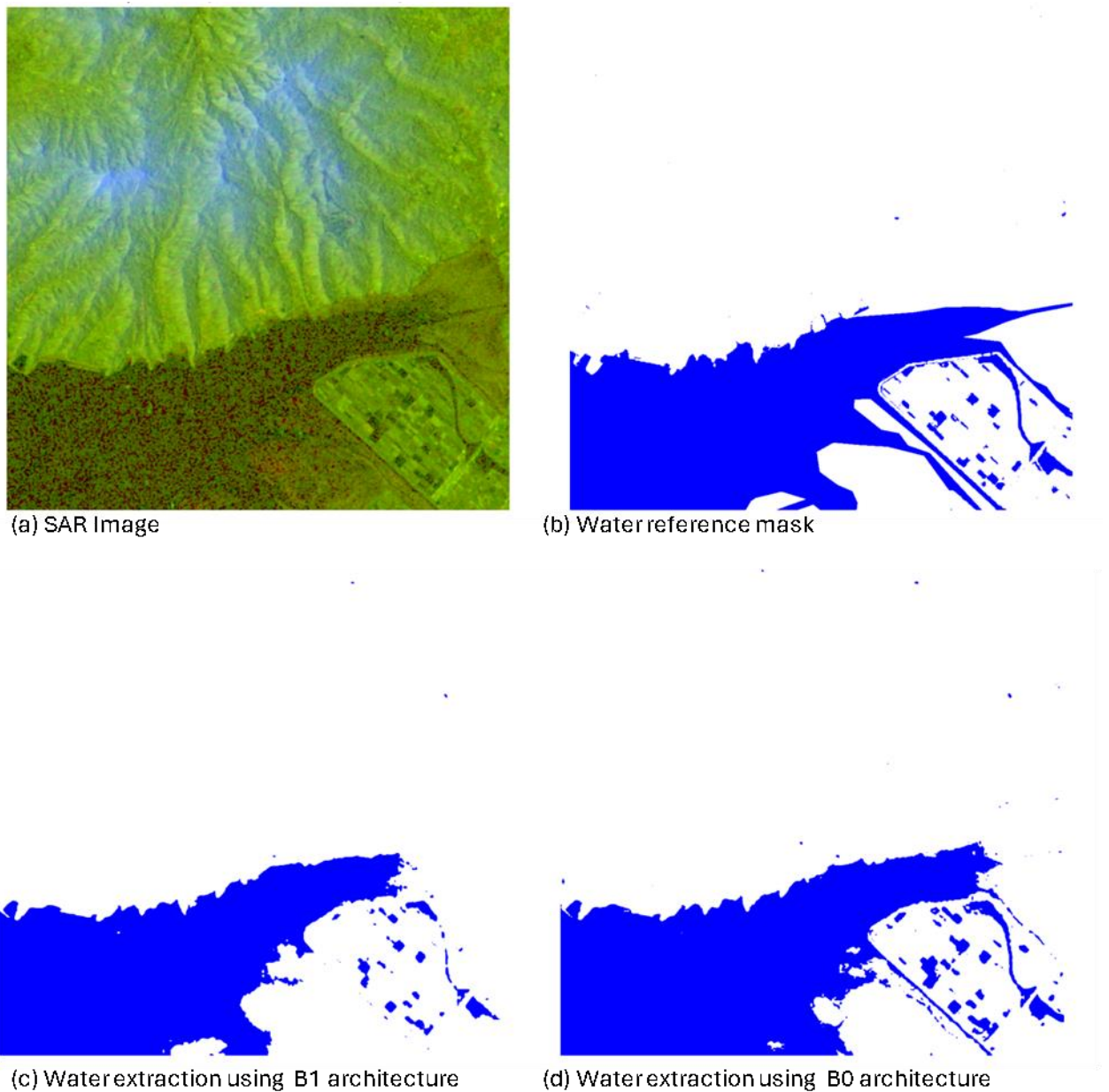
(a) SAR Image

(b) Water reference mask

(c) Water extraction using B1 architecture

(d) Water extraction using B0 architecture

Figure 6 Issues in complex (smooth) water-land transition zones

Conclusion

This study developed and validated a robust, multi-modal deep learning framework for water segmentation. The central finding is that for this task, increased model capacity does not necessarily yield superior performance for water segmentation. Our comparative analysis demonstrated that both the lightweight SegFormer-B0 model and larger B1 model consistently achieved similar accuracy (Water IoU of 0.9582 and 0.9560, respectively). Critically, this accuracy was achieved by B0 model with a ~48% faster inference speed, establishing the B0 architecture as the optimal choice. This suggests that the feature space for distinguishing water from non-water using a fusion of SAR and topographic data can be fully captured by the representational power of the MiT-B0

backbone, while the additional parameters of the B1 model offer no significant advantage.

The framework's methodological contributions further enhance its operational value. The proposed per-scene normalization strategy was proven effective, enabling the model to generalize successfully across different C-band SAR sensors (Sentinel-1 and EOS-04) and radiometric products (Sigma Naught, Beta Naught & Gamma Naught. Qualitatively, the model demonstrated a strong ability to handle traditionally challenging scenarios, leveraging the transformer's global context and the fused topographic data to delineate small waterbodies and river networks and correctly classify water bodies even when confounded by thin ice cover. While a multi-pass, overlapping inference strategy was evaluated and shown to provide a marginal increase in IoU, our results indicate that a single-pass approach is largely sufficient, reinforcing the framework's efficiency.

Despite its high overall performance, the model's limitations in complex water-land transition zones highlight a clear avenue for future research. Nevertheless, the presented SegFormer-B0 framework provides a highly accurate, computationally efficient, and robust solution for large-scale, operational water body monitoring.

References

1. D. Gupta, V. Kushwaha, A. Gupta and P. K. Singh, "Deep Learning based Detection of Water Bodies using Satellite Images," 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, 2021, pp. 1-4, doi: 10.1109/CONIT51480.2021.9498442

2. Druce, D., Tong, X., Lei, X., Guo, T., Kittel, C. M., Grogan, K., & Tottrup, C. (2021). An optical and SAR based fusion approach for mapping surface water dynamics over mainland China. *Remote Sensing, 13*(9), 1663.

3. Wang, Y.; Li, S.; Lin, Y.; Wang, M. Lightweight Deep Neural Network Method for Water Body Extraction from High-Resolution Remote Sensing Images with Multisensors. *Sensors* **2021**, *21*, 7397.

4. Nagaraj, R.; Kumar, L.S. Extraction of Surface Water Bodies Using Optical Remote Sensing Images: A Review. *Earth Sci Inf.* **2024**, *17*, 893–956.

5. Kadapala, B.K.R., Hakeem, K.A., Cheruvupally, S.K. *et al*. A Connected Pixels Based Approach for Surface Water Layer Extraction from EOS-04 Medium Resolution ScanSAR (MRS) Data. *J Indian Soc Remote Sens* **52**, 2727–2737 (2024). https://doi.org/10.1007/s12524-024-02018-6

6. B. K. R. Kadapala and A. H. K, "Region-Growing-Based Automatic Localized Adaptive Thresholding Algorithm for Water Extraction Using Sentinel-2 MSI Imagery," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1-8, 2023, Art no. 4201708, doi: 10.1109/TGRS.2023.3246540.

7. Hakimdavar, R., Hubbard, A., Policelli, F., Pickens, A., Hansen, M., Fatoyinbo, T., Lagomasino, D., Pahlevan, N., Unninayar, S., Kavvada, A., Carroll, M., Smith, B., Hurwitz, M., Wood, D., & Schollaert Uz, S. (2020). Monitoring Water-Related Ecosystems with Earth Observation Data in Support of Sustainable Development Goal (SDG) 6 Reporting. *Remote Sensing*, *12*(10), 1634. https://doi.org/10.3390/rs12101634

8. Palmer, Stephanie CJ, Tiit Kutser, and Peter D. Hunter. "Remote sensing of inland waters: Challenges, progress and future directions." *Remote sensing of Environment* 157 (2015): 1-8.

9. Meng, L.; Guo, S.; Li, S. Summary on Extraction of Water Body from Remote Sensing Image and Flood Monitoring. *Water Conserv. Informatiz*. **2012**

10. Woolway, R.I.; Kraemer, B.M.; Lenters, J.D.; Merchant, C.J.; O'Reilly, C.M.; Sharma, S. Global lake responses to climate change. *Nat. Rev. Earth Environ.* **2020**, *1*, 388–403.

11. Deke YAO, Guoqiang JIA, Yubao QIU, A dataset of annual surface water extent in the Xijiang River Basin from 1986 to 2023, China Scientific Data, 10.11922/11-6035.csd.2025.0014.zh, **10**, 3, (1-20), (2025).

12. Kaplan, G., & Avdan, U. (2018). Monthly analysis of wetlands dynamics using remote sensing data. *ISPRS International Journal of Geo-Information*, *7*(10), 411.

13. Dan, L. I., Baosheng, W. U., Bowei, C. H. E. N., Yuan, X. U. E., & Yi, Z. H. A. N. G. (2020). Review of water body information extraction based on satellite remote sensing. *Journal of Tsinghua University (Science and Technology)*, *60*(2), 147-161.

14. Pekel, JF., Cottam, A., Gorelick, N. *et al*. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422 (2016). https://doi.org/10.1038/nature20584

15. Chen, F. Comparing Methods for Segmenting Supra-Glacial Lakes and Surface Features in the Mount Everest Region of the Himalayas Using Chinese GaoFen-3 SAR Images. *Remote Sens*. **2021**, *13*, 2429.

16. Silveira, M.; Heleno, S. Water/land segmentation in SAR images using level sets. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 1896–1899.

17. Meng, Q.; Wen, X.; Yuan, L.; Xu, H. Factorization-Based Active Contour for Water-Land SAR Image Segmentation via the Fusion of Features. *IEEE Access* **2019**, *7*, 40347–40358.

18. Zheng, Y.; Zhang, X.; Hou, B.; Liu, G. Using Combined Difference Image and k - Means Clustering for SAR Image Change Detection. *IEEE Geosci. Remote Sens. Lett*. **2014**, *11*, 691–695.

19. Liu, Z.; Li, F.; Li, N.; Wang, R.; Zhang, H. A Novel Region-Merging Approach for Coastline Extraction from Sentinel-1A IW Mode SAR Imagery. *IEEE Geosci. Remote Sens. Lett*. **2016**, *13*, 324–328.

20. Li, Ning, et al. "Dynamic waterline mapping of inland great lakes using time-series SAR data from GF-3 and S-1A satellites: A case study of DJK reservoir, China." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12.11 (2019): 4297-4314.

21. Lv, W.; Yu, Q.; Yu, W. Water extraction in SAR images using GLCM and Support Vector Machine. In Proceedings of the IEEE 10th International Conference on Signal Processing Proceedings, Beijing, China, 24–28 October 2010; pp. 740–743.

22. Cafaro, B., Canale, S., & Pirri, F. (2012, July). X-SAR SpotLigh images feature selection and water segmentation. In *2012 IEEE International Conference on Imaging Systems and Techniques Proceedings* (pp. 217-222). IEEE.

23. Guo, Z., Wu, L., Huang, Y., Guo, Z., Zhao, J., & Li, N. (2022). Water-Body Segmentation for SAR Images: Past, Current, and Future. *Remote Sensing*, *14*(7), 1752. https://doi.org/10.3390/rs14071752

24. Ren, Y.; Xu, H.; Liu, B.; Li, X. Sea Ice and Open Water Classification of SAR Images Using a Deep Learning Model. In Proceedings of the IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 3051–3054.

25. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

26. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Lecture Notes in Computer Science*; Springer International Publishing: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

27. Katiyar, V., Tamkuan, N., & Nagai, M. (2021). Near-real-time flood mapping using off-the-shelf models with SAR imagery and deep learning. *Remote Sensing*, *13*(12), 2334.

28. Konapala, G., Kumar, S. V., & Ahmad, S. K. (2021). Exploring Sentinel-1 and Sentinel-2 diversity for flood inundation mapping using deep learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *180*, 163-173.

29. Bai, X., Zhang, Y., & Wei, J. (2025). LGFUNet: A Water Extraction Network in SAR Images Based on Multiscale Local Features with Global Information. *Sensors*, *25*(12), 3814. https://doi.org/10.3390/s25123814

30. Ren, Y.; Li, X.; Yang, X.; Xu, H. Development of a Dual-Attention U-Net Model for Sea Ice and Open Water Classification on SAR Images. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5.

31. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the Computer Vision–ECCV 2018, Munich, Germany, 6 October 2018; pp. 833–851.

32. Kim, M. U., Oh, H., Lee, S. J., Choi, Y., & Han, S. (2021, July). Deep learning based water segmentation using KOMPSAT-5 SAR images. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (pp. 4055-4058). IEEE.

33. Ding, L., Lin, D., Lin, S., Zhang, J., Cui, X., Wang, Y., ... & Bruzzone, L. (2022). Looking outside the window: Wide-context transformer for the semantic segmentation of high-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-13.

34. Wang, L., Fang, S., Meng, X., & Li, R. (2022). Building extraction with vision transformer. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1-11.

35. Arora, V., Gupta, S., Kudupu, A., Priyadarshi, A., Mundayatt, A., & Sreevalsan-Nair, J. (2025). CCESAR: Coastline Classification-Extraction From SAR Images Using CNN-U-Net Combination. arXiv preprint arXiv:2501.12384.

36. Lin, X., Cheng, Y., Chen, G., Chen, W., Chen, R., Gao, D., Zhang, Y., & Wu, Y. (2023). Semantic Segmentation of China's Coastal Wetlands Based on Sentinel-2 and Segformer. *Remote Sensing*, *15*(15), 3714. https://doi.org/10.3390/rs15153714

37. Li, M., Rui, J., Yang, S., Liu, Z., Ren, L., Ma, L., Li, Q., Su, X., & Zuo, X. (2023). Method of Building Detection in Optical Remote Sensing Images Based on SegFormer. *Sensors*, *23*(3), 1258. https://doi.org/10.3390/s23031258

38. Wieland, M., Fichtner, F., Martinis, S., Groth, S., Krullikowski, C., Plank, S., Motagh, M. (2023). S1S2-Water: A global dataset for semantic segmentation of water bodies from Sentinel-1 and Sentinel-2 satellite images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, doi: 10.1109/JSTARS.2023.3333969.