Water Region Extraction in Aerial Images Using Deep Learning

Nikita Pandey
Dept. of Computer Science and Engineering
Pranveer Singh Institute of Technology
Kanpur, India
nikitapandey2803@gmail.com

Saumya Kumar
Dept. of Computer Science and Engineering
Pranveer Singh Institute of Technology
Kanpur, India
saumyakumar.09@gmail.com

Satish Kumar Yadav
Dept. of Computer Science and Engineering
Pranveer Singh Institute of Technology
Kanpur, India
satish34557@gmail.com

Shashwat Srivastava
Dept. of Computer Science and Engineering
Pranveer Singh Institute of Technology
Kanpur, India
29800csaiml@gmail.com

Shomil Singh
Dept. of Computer Science and Engineering
Pranveer Singh Institute of Technology
Kanpur, India
shomil02020@gmail.com

Deepak Gupta
Dept. of Computer Science and Engineering
Pranveer Singh Institute of Technology
Kanpur, India
deepakgupta.cse6@gmail.com

Abstract— Water bodies are essential component of Earth's surface and accurate water region mapping is crucial for disaster management and environmental monitoring. Existing methods to map water bodies relies on traditional modelling methods that need extensive manual calibration struggle with scalability and lack advanced pattern recognition. This paper evaluates the effectiveness of deep learning-based semantic segmentation models, DeepLabV3+ and U-Net, using aerial images resized to 256x256 pixels in mapping water regions in aerial images. DeepLabV3+, with a ResNet50 backbone, and U-Net are trained on a comprehensive flood-affected dataset with preprocessing techniques such as resizing and augmentation. Model performance is assessed using metrics like IoU, F1 score, accuracy, precision, and recall. The results indicate that DeepLabV3+ outperforms U-Net, showing promise for automating water region mapping and improving flood surveys and urban planning.

Keywords— Water body mapping, semantic segmentation, DeepLabV3+, U-Net, disaster Management.

I. INTRODUCTION

Water resources such as oceans, rivers, lakes, and reservoirs are essential for sustaining ecosystems, but they are increasingly threatened by rapid urbanization [1]. Flooding, one of the most frequent and catastrophic natural disasters, accounted for 43% of all recorded incidents globally between 1998 and 2017, affecting over two billion people (UNISDR, 2017) [2]. As populations grow and economic activity increases in flood-prone areas, the impact of floods continues to rise. To effectively manage and preserve water resources, accurate surveying and continuous monitoring are crucial [3].

For decades, Landsat satellite imagery has been used to map surface water extent by analyzing water's spectral signature. In recent years, the advent of high-resolution remote sensing data from commercial satellites has enhanced our ability to monitor surface water [2]. These advancements, coupled with the high spatial resolution, broad coverage, and short revisit intervals of satellite-based earth observation techniques, have made them increasingly popular for surface water monitoring [4]. However, extracting detailed information from high-resolution images remains challenging due to the complex attributes of water bodies [5]. This study compares two advanced deep learning models, U-Net and DeepLabV3+, commonly used for semantic segmentation. U-Net's encoder-decoder architecture effectively captures local and global context, while DeepLabV3+ uses atrous convolution and spatial pyramid pooling to handle complex segmentation tasks. This study aims to evaluate their performance, offering insights for improving the accuracy and efficiency of surface water detection and management.

II. LITERATURE REVIEW

Water bodies are a vital component of Earth's ecosystem, accounting for only 1.75% of the planet's total water storage. Surface water is crucial for maintaining biodiversity, influencing climate change, and sustaining the global water cycle [6]. However, rapid industrial development and human activity have increasingly led to surface water pollution and depletion. Therefore, accurately identifying, and delineating surface water bodies is essential for promoting sustainable development, environmental conservation, and effective

979-8-3503-6917-5/24/$31.00 ©2024 IEEE

urban planning. Even while manual surveys are dependable, they can be time- and labor-intensive. Remote sensing has been an effective method for addressing these issues, particularly in urban hydrological studies that try to enhance and manage urban water systems [7]. Mapping aquatic bodies has become more dependent on advances in satellite imagery, particularly high-resolution photos, and advanced image processing methods. Planning for efficient flood protection and water quality management requires access to extensive information about the dynamics, distribution, and quality of water bodies, which is provided by these technologies [8].

One popular use of remote sensing has been the mapping of surface water [9]. Recent developments in remote sensing have made it possible to exploit data collected in real time by a variety of sensors on board satellites [1]. Many studies throughout the world have shown that satellite photography is an effective tool for mapping aquatic bodies. Understanding changes in the water cycle associated with global warming and evaluating the availability of water resources depend on semantic segmentation, or the identification of water bodies from satellite pictures. Observations from remote sensing are also helpful for monitoring floods [10]. Semantic picture segmentation has several uses, from robotics and medical image analysis [11].

Deep convolutional neural networks (DCNNs) have significantly enhanced water body detection accuracy. Models such as Fully Convolutional Networks, U-Net, and DeepLabV3+ offer considerable improvements over traditional methods by allowing for more precise segmentation and analysis of water bodies [11, 12]. Semantic segmentation using these models has proven effective in various remote sensing applications, including assessing water resources and monitoring flood events. However, challenges persist in accurately detecting small rivers, paddy fields, and complex riverbeds. These difficulties highlight the need for more advanced and efficient deep learning network structures [5]. Additionally, the performance of models like DeepLabV3+ is heavily reliant on the quality and diversity of training data, underscoring the importance of continuous advancements in data collection and processing to improve model accuracy and reliability [13].

III. MATERIALS AND METHODS

A. Dataset Overview

The "Flood Area Segmentation Dataset" [14] is designed for water region segmentation in flood-affected areas. It consists of 290 high-resolution aerial images (see fig.-1), each paired with a mask image that delineates water regions. These images cover a wide range of flood scenarios across different geographical and urban landscapes. The mask images, created using Label Studio, provide high-quality, precise segmentation with water regions in white and other areas in black. To maintain consistency and manage computational resources, all images and masks were resized to 256x256 pixels. The dataset was divided into training, validation, and test sets using an 8:1:1 ratio to optimize model training, hyperparameter tuning, and performance evaluation. To expand the training dataset and enhance model robustness, augmentation techniques were applied, including rotations at 45° intervals, effectively generating 8 distinct orientations for each image.

(a) Image (b) Ground Truth

Fig. 1. Image and its mask

This method ensures the model's resilience to directional variations by systematically covering all major orientations.

B. Methodology

The proposed methodology (Figure 2) uses Semantic Segmentation for identifying water regions from provided aerial images. Semantic segmentation methods based on deep learning can map pixels to semantics and mine spectral features of deep remote sensing images [15]. It classifies each pixel in an image into a predefined category. In this process, the goal is to label every pixel with the class of the object it belongs to, that is to label pixels covering the water region and other areas. Semantic segmentation gives us a more comprehensive, pixel-level understanding of the image, enabling us to get more precise insights and efficient analysis of images [16]. For Semantic Segmentation we are using two deep learning models: U-Net and DeepLabV3+ Model.

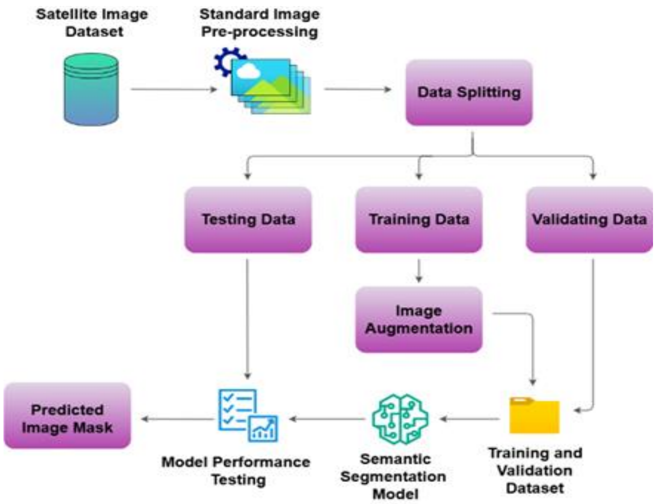


Fig. 2. Proposed Methodology

U-Net [17] is designed for semantic segmentation and features a symmetrical encoder-decoder design. The encoder captures detailed features through downsampling, while the decoder reconstructs the image using upsampling and incorporates these features through skip connections. These connections improve the precision of segmentation by linking parts of the encoder and decoder. For flood water mapping, U-Net excels in providing detailed, accurate pixel-level classifications.

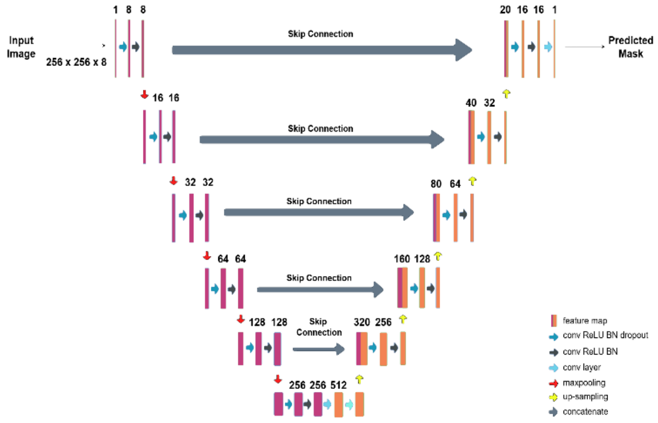


Fig. 3. U-Net Architecture

DeepLabV3+ [18] uses atrous convolutions and spatial pyramid pooling to handle complex segmentation tasks. Its ResNet50-based encoder extracts deep features, and the decoder applies spatial pyramid pooling to keep high-resolution details intact. While it does not use skip connections, it excels at combining contextual information across different scales. In flood mapping, DeepLabV3+ provides precise segmentation with detail and effective multi-scale context [19].

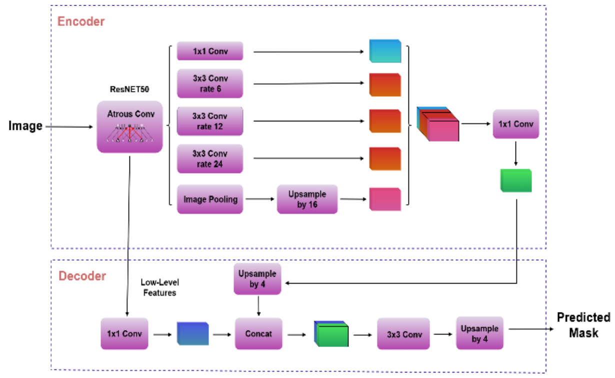


Fig. 4. DeepLabV3+ Architecture

C. Image Resizing

Image resizing involves altering input dimensions to meet model requirements, typically by scaling, cropping, or padding to achieve a specific size (figure 5). Variable-sized images were resized to a uniform 256x256 pixels to align with the architectural needs of U-Net and DeepLabv3+ models. U-Net, with its symmetric expansion path, requires consistent input dimensions to merge high-resolution encoder features with upsampled decoder features, making sizes like 256x256 ideal due to its downsampling and upsampling structure, ensuring precise segmentation [20]. Similarly, DeepLabv3+ employs atrous convolutions to capture multi-scale context while preserving spatial resolution, and the standardized 256x256 size allows these convolutions to effectively maintain detail and accuracy. Uniform image sizes also support efficient batch processing and optimal GPU utilization.

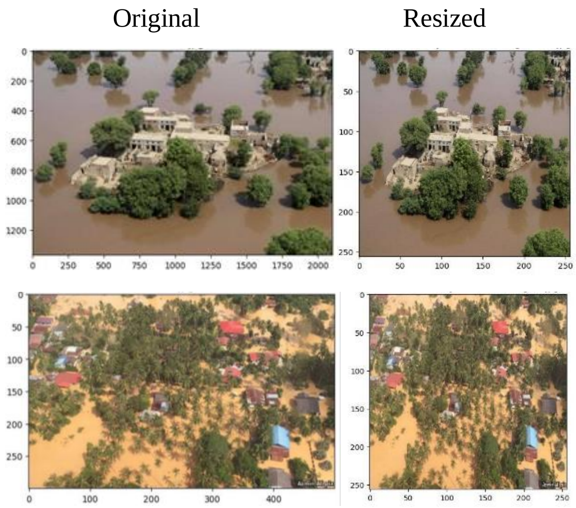


Fig. 5. Image Resizing

D. Image Augmentation

Image augmentation increases the training dataset volume by applying transformations such as rotations, flips, scaling, and colour adjustments. These techniques enhance data diversity and model robustness. For semantic segmentation models like U-Net and DeepLabV3+, augmentation improves generalization by exposing the network to various image variations and addressing class imbalance. Padding ensures that augmented images conform to the 256x256 pixel standard required by the model (see fig.-6).

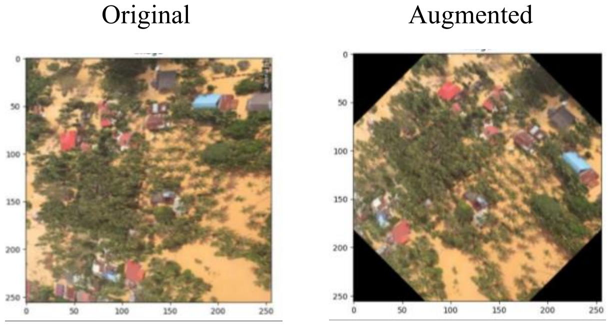


Fig. 6. Image Augmentation

E. Model Training and Testing

U-Net and DeepLabv3+ models are trained using training and validation dataset. After training, the models are tested with the help of test dataset. Model produces image masks of flooded areas which are then compared with the masks in the given dataset to generate evaluation metrics for model's performance comparison.

IV. RESULTS AND DISCUSSION

A comparative performance analysis of UNET and DeepLabV3+, two cutting-edge deep learning models for water region detection is describes in table-1. The performance of each model is assessed based on evaluation metrics, including training accuracy and loss, Intersection over Union (IoU), precision, recall, and F1 score.

A. Accuracy: Training accuracy gauges the ratio of correctly classified samples relative to the total number of samples during the training process.

$$\text{Accuracy} = \frac{\text{Total number of CCP}}{\text{Total number of instances}} \quad (1)$$

B. Training Loss: Binary cross-entropy (BCE) function is used to calculate loss while training the model.

$$\text{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[a_i \log(b_i) + (1 - a_i)\log b_i] \quad (2)$$

where, a_i - denotes actual ground truth value for i^{th} pixel
b_i - denotes predicted confidence value for i^{th} pixel

C. Intersection over Union (IoU): IoU evaluates the degree of intersection between the predicted and ground truth regions, providing an assessment of segmentation quality.

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (3)$$

D. Precision: Precision assesses how accurately the model's positive predictions align with the actual positive instances.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

where, TP – True positives
FP – False Positives

E. Recall: Recall gauges the model's effectiveness in detecting all relevant positive instances.

$$Recall = \frac{TP}{TP + TN} \quad (5)$$

where, TP – True Positives
TN – True Negatives

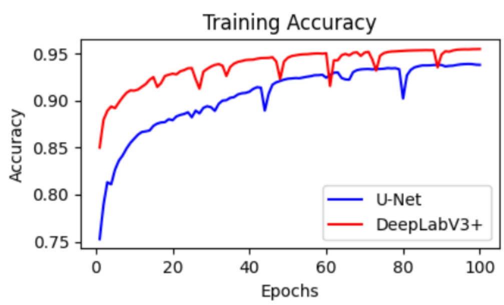
F. F1 Score: The F1 score balances precision and recall, providing a comprehensive evaluation of the model's performance by accounting for both false positives and false negatives.

$$F1\ Score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (6)$$

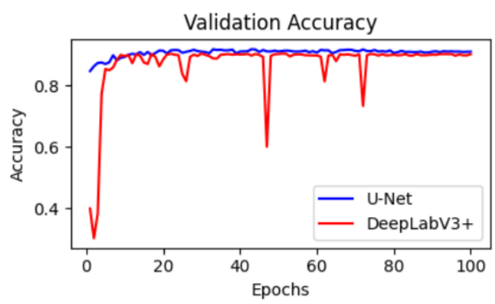
Based on results obtained, it is observed that the models stabilized at 100 epochs and provided best results. Under the same training conditions, the UNET model attained a training accuracy of 93.93% with a training loss of 8.93% for 100 epochs. It also recorded an IoU of 89.77%, an F1 score of 95.34%, precision of 95.40%, and recall of 95.28%. In comparison, the DeepLabV3+ model outperformed UNET by attaining a training accuracy of 95.47% and a lower training loss of 4.10%. It achieved an IoU of 95.27%, an F1 score of 97.94%, precision of 97.95%, and recall of 97.93%. These results indicate that DeepLabV3+ is more accurate and efficient in segmenting water regions, consistently providing higher evaluation metrics and lower loss values compared to the UNET model.

Table 1. Model Performance

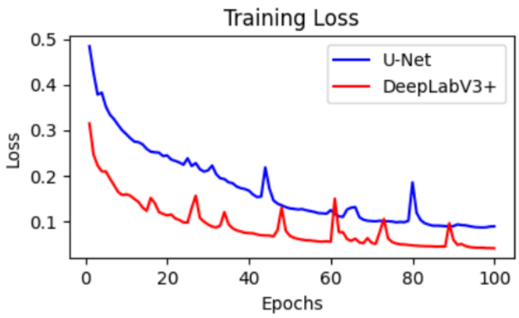
Metric	Epochs	U-Net	DeeplabV3+
Accuracy	50	0.9101	0.9453
	75	0.9165	0.9524
	100	0.9393	0.9547
Loss	50	0.1368	0.063
	75	0.1091	0.0472
	100	0.077	0.041
IOU	50	0.8179	0.9162
	75	0.8349	0.9428
	100	0.8977	0.9527
Precision	50	0.9305	0.9651
	75	0.9417	9,749
	100	0.9587	0.9795
Recall	50	0.9214	0.965
	75	0.9396	0.9746
	100	0.9584	0.9794
F1 Score	50	0.9124	0.9625
	75	0.9271	0.9748
	100	0.9534	0.9794



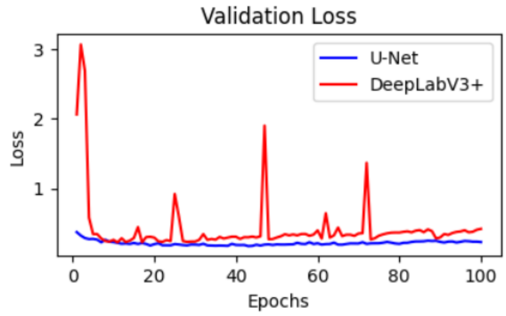
(a) Training Accuracy



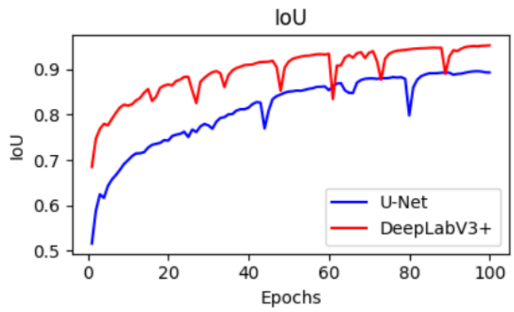
(b) Validation Accuracy



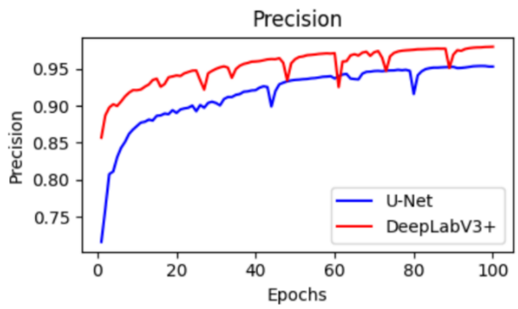
(c) Training Loss



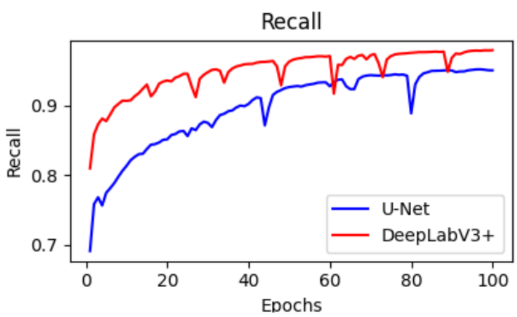
(d) Validation Loss



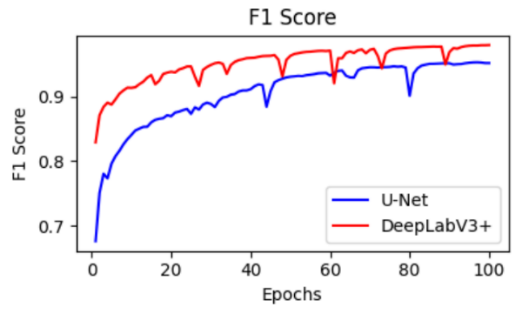
(e) IoU



(f) Precision



(g) Recall



(h) F1-Score

Fig. 7. Performance Analysis- (a) Training Accuracy (b)Validation Accuracy (c) Training Loss (d) Validation Loss (e) IoU (f) Precision (g) Recall (f) F1-Score

The comparative analysis is shown in Figure 7. It is recorded that upon increasing the epochs value model generalizes itself and can learn the feature more effectively. The continuous growth in accuracy, IoU and F1 score makes the model robust and make it a suitable choice for water body extraction from aerial images.

V. CONCLUSION

In this study, two advanced deep learning models, UNET and DeepLabV3+, were evaluated for the task of water region detection. The investigation involved training both models under identical conditions for 50, 75, and 100 epochs, with a focus on identifying the optimal training duration. The analysis demonstrated that DeepLabV3+ consistently outperformed UNET across various evaluation metrics, including training accuracy, Intersection over Union (IoU), F1 score, precision, and recall.

The DeepLabV3+ model demonstrated exceptional performance with a training accuracy of 95.47%, IoU of 95.27%, F1 score of 97.94%, precision of 97.95%, and recall of 97.93%, coupled with lower training loss. In comparison, the U-Net model, although effective, showed lower accuracy and higher loss values. These results indicate that DeepLabV3+ is more adept at accurately segmenting water regions due to its advanced architecture, making it the preferred choice for this task. However, model selection may vary based on specific needs and resource constraints; U-Net may be suitable for scenarios with limited computational resources, while DeepLabV3+ is ideal for cases where maximizing accuracy is crucial. This study offers valuable insights for future research and applications in water region detection, guiding the development of more refined models and techniques in this field.

VI. FUTURE SCOPE

As the current study focuses on flood mapping, the techniques developed here have potential applications in various fields related to water region monitoring and management. Future research could explore the following areas: Coastal Zone Management, Wetland Conservation, Reservoir and Dam Management, River System Dynamics, Glacier and Snow Melt Monitoring, Agricultural Water Management, Urban Water Infrastructure and many more. To improve model robustness and generalization, techniques like domain adaptation, few-shot learning, and adversarial training could be investigated.

REFERENCES

[1] Rishikesh G. Tambe, Sanjay N. Talbar, Satishkumar S. Chavan, “Deep multi-feature learning architecture for water body segmentation from satellite images”, Journal of Visual Communication and Image Representation, 77, p.103141, 2021.

[2] Shastry, Apoorva, Elizabeth Carter, Brian Coltin, Rachel Sleeter, Scott McMichael, and Jack Eggleston. "Mapping floods from remote sensing data and quantifying the effects of surface obstruction by clouds and vegetation", Remote Sensing of Environment 291, p.113556, 2023.

[3] Pech-May F, Aquino-Santos R, Delgadillo-Partida J. “Sentinel-1 SAR images and deep learning for water body mapping”, Remote Sensing *15*(12), p.3009, 2023.

[4] Junjie Li, Linyi Li, Yanjiao Song, Jiaming Chen, Zhe Wang, Yi Bao, Wen Zhang, Lingkui Meng. “A robust large-scale surface water mapping framework with high spatiotemporal resolution based on the fusion of multi-source remote sensing data,” International Journal of Applied Earth Observation and Geoinformation, p.103288, 2023.

[5] Sun, Dechao, Guang Gao, Lijun Huang, Yunpeng Liu, and Dongquan Liu, "Extraction of water bodies from high-resolution remote sensing imagery based on a deep semantic segmentation network.", Scientific Reports 14, no. 1, 14604, 2024.

[6] Wu, Peng, et al, "Research on water extraction from high resolution remote sensing images based on deep learning", Frontiers in Remote Sensing 4, p.1283615, 2023.

[7] Wang, Xianwei, and Hongjie Xie, "A review on applications of remote sensing and geographic information systems (GIS) in water resources and flood risk management," *Water 10*(5), p.608, 2018.

[8] Reham Gharbia, “Deep Learning for Automatic Extraction of Water Bodies Using Satellite Imagery”, in Journal of the Indian Society of Remote Sensing, *51*(7), pp.1511-1521, 2023

[9] Mayer, Timothy, et al, "Deep learning approach for Sentinel-1 surface water mapping leveraging Google Earth Engine," ISPRS Open Journal of Photogrammetry and Remote Sensing 2, p.100005, 2021.

[10] Mina Talal, Alavikunhu Panthakkan, Husameldin Mukhtar, Wathiq Mansoor, Saeed Almansoori‡ and Hussain Al Ahmad, “Detection of Water-Bodies Using Semantic Segmentation”, in International Conference on Signal Processing and Information Security (ICSPIS), pp. 1-4, IEEE, 2018

[11] Edpuganti, A., Akshaya, P., Gouthami, J., Sajith Variyar, V.V., Sowmya, V. and Sivanpillai, R., 2023. “Effect of data quality on water body segmentation with deeplabv3+ algorithm”. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, pp.81-85.

[12] Yuan, Kunhao, et al, "Deep-learning-based multispectral satellite image segmentation for water body detection", in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, pp.7422-7434, 2021

[13] George, Bichu, et al., "Performance Improvement of Water Body Segmentation by DeeplabV3+ Using Two-Dimensional Variational Mode Decomposition", in 10th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 603-608, IEEE, 2023.

[14] Dataset:https://www.kaggle.com/datasets/faizalkarim/flood-area-segmentation/data

[15] Ziyao Li, Rui Wang, Wen Zhang, Fengmin Hu, And Lingkui Meng, “Multiscale Features Supported DeepLabV3+ Optimization Scheme for Accurate Water Semantic Segmentation”, supported by the National Key Research and Development Program of China Vol-7, IEEE Access, 7, pp.155787-155804, 2019

[16] Badrinarayanan V, Kendall A, Cipolla R, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE transactions on pattern analysis and machine intelligence, *39*(12), pp.2481-2495, 2017.

[17] Ling, Gui, et al, "Real-time water area segmentation for USV using enhanced u-net", 2020 Chinese Automation Congress (CAC), IEEE, pp. 2533-2538, 2020.

[18] Harika, A., et al, "Extracting water bodies in RGB images using deeplabv3+ algorithm", in The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 46, pp.97-101, (2022).

[19] Zeng H, Peng S, Li D, “Deeplabv3+ semantic segmentation model based on feature cross attention mechanism”, in Journal of Physics: Conference Series Vol. 1678, No. 1, p. 012106, 2020.

[20] Özelkan E., “Water body detection analysis using NDWI indices derived from landsat-8 OLI”, in Polish Journal of Environmental Studies, *29*(2), pp.1759-1769, 2020.