Segmentation of River Networks from Multispectral Remote Sensing Data Using Deep Neural Networks

Eduardo H. P. Souza[1], **Francisco de A. Boldt**[1], **Thiago M. Paixão**[1], **Jefferson O. Andrade**[1], **Karin S. Komati**[1]

[1]Programa de Pós-graduação em Computação Aplicada (PPComp)
Instituto Federal do Espirito Santo (IFES), Campus Serra
Av. dos Sabiás, 330 - Morada de Laranjeiras, Serra - ES, Brazil, 29166-630

duvrdx@gmail.com

{franciscoa,thiago.paixao,jefferson.oliveira,kkomati}@ifes.edu.br

Abstract. Research Context: *Accurate hydrological mapping is essential for water resource management, flow modeling, and risk assessment of flooding and erosion.* ***Scientific Problem:*** *Conventional methods and RGB satellite imagery fail to map watercourses obscured by dense vegetation, resulting in fragmented, unreliable cartography that compromises territorial planning.* ***Proposed Solution:*** *We propose a comparative analysis of river network semantic segmentation, evaluating U-Net and DeepLabv3+ architectures with ImageNet pre-trained backbones (EfficientNet-B5, ResNet-152). The analysis utilizes multispectral data (RGB, NIR) and spectral indices (NDWI, NDVI, GNDVI) to enhance feature discrimination.* ***Related IS Theory:*** *This work aligns with Task-Technology Fit by evaluating the suitability of deep learning for automated river network extraction.* ***Research Method:*** *We created a dataset from satellite imagery and official vector data of two Brazilian basins (Doce, Itapemirim). Models were trained in intra-dataset, combined, and cross-dataset scenarios using mean Intersection over Union (mIoU).* ***Summary of Results:*** *For intra-dataset evaluation, U-Net using EfficientNet-B5 yielded 84.86±0.39% mIoU, whereas DeepLabv3+ with ResNet-152 showed 86.73±0.21%. Combined datasets provided stable metrics (82.25±0.51% for U-Net). Cross-dataset tests reached 63.96% mIoU, indicating a performance decrease.* ***Contributions/Impact:*** *This study demonstrates that U-Net and DeepLabv3+ architectures using deep backbones accurately segment river networks in complex environments. The methodology provides a viable approach to automate drainage network extraction, supporting environmental management and hydrological modeling.*

1. Introduction

Mapping of hydrographic networks supports the management of water resources in countries with large river systems, such as Brazil [Viana et al. 2024]. However, conventional methods based on remote sensing images face a challenge caused by the occlusion of watercourses under riparian vegetation [Rusnák et al. 2022]. RGB satellite images have limitations, including their inability to capture spectral information beyond the visible range, which makes it challenging to distinguish complex phenomena. They may also fail to capture fine details of small water bodies or irregular edges and are unable to detect vegetation occlusion, as shown in the results of Sun et al. [Sun et al. 2024]. This phenomenon

often leads to fragmented and unreliable cartographic products, directly compromising territorial planning and environmental monitoring.

To overcome these limitations, multispectral imagery provides a broader range of spectral information, which can enhance the identification of surface features. As noted by Yuan et al. [Yuan et al. 2021], the effective use of broader spectral bands from multispectral sensors to achieve improved performance over RGB bands remains underexplored. Addressing this gap, the present work incorporates spectral components beyond the visible range, particularly the Near-Infrared (NIR) band, and employs derived indices such as the Normalized Difference Vegetation Index (NDVI), the Green Normalized Difference Vegetation Index (GNDVI), and the Normalized Difference Water Index (NDWI). These features are expected to enhance the detection and characterization of water bodies under vegetation occlusion.

While multispectral and index-based approaches can enhance spectral discrimination, segmenting water bodies in complex environments remains a challenge. Employing deep learning techniques for the semantic segmentation of remote sensing imagery has become a common strategy for extracting surface water information [Cao et al. 2024]. This study investigates the use of two prominent convolutional neural network architectures: U-Net [Ronneberger et al. 2015] and DeepLabv3+ [Chen et al. 2018]. The U-Net is widely recognized for its encoder–decoder structure, which effectively integrates contextual and spatial information to delineate precise boundaries. Complementing this, DeepLabv3+ employs Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale context, a critical feature for identifying continuous river segments of varying widths.

To improve segmentation performance, we apply transfer learning by incorporating pre-trained convolutional neural networks (CNNs) as feature extraction backbones for both architectures. Models such as ResNet-152 [He et al. 2016] and EfficientNet [Tan and Le 2019], trained on the ImageNet dataset [Fei-Fei et al. 2009], provide hierarchical features learned from a large number of natural images. This approach enables the networks to detect textures and spatial patterns that indicate the presence of watercourses beneath vegetation, even when the training data are limited.

Model performance was quantitatively evaluated using standard segmentation metrics, including Intersection over Union (IoU) and mean IoU (mIoU). The study utilized two distinct labeled datasets corresponding to the Itapemirim and Doce river basins, developed from official hydrographic data and satellite imagery from Espírito Santo. The experimental framework comprised three scenarios: intra-dataset evaluation, combined dataset training, and cross-dataset generalization tests.

In summary, the main contributions of this work are as follows:

- Two separate multispectral datasets were developed for river network segmentation, each corresponding to a specific Brazilian river basin in Espírito Santo: the Itapemirim and the Doce River. Both were created by integrating satellite imagery with official hydrographic vector data.
- The use of multispectral and index-based features within U-Net and DeepLabv3+ architectures employing ResNet-152 and EfficientNet-B5 backbones, to assess how additional spectral information influences the performance of hydrographic segmentation under vegetation occlusion.

- An evaluation of the models' generalization capability through cross-dataset testing, providing evidence of their applicability across different geographical regions.

This article is organized as follows: Section 2 presents related work. Section 3 describes the materials and methods, including the U-Net architecture and the chosen backbones. The experimental results are presented and discussed in Section 4. Finally, Section 5 draws the conclusions and suggests directions for future work.

2. Related Work

Machine learning models, including Random Forest, Support Vector Machine, and K-Nearest Neighbours, were also developed for Land Use and Land Cover classification using remote sensing data in the Quadrilátero Ferrífero region of Brazil [Osias et al. 2024]. In this work, the models were trained to classify imagery into seven LULC classes (Rocky Outcrop, Low Vegetation, Mining, Urban Area, Eucalyptus, Forest, and Rivers and Lakes). When compared, the models demonstrated no significant difference in performance, achieving F1-scores ranging from approximately 0.89 to 0.93 and Kappa coefficients from 0.88 to 0.92.

The U-Net architecture has established itself as a standard baseline for semantic segmentation across diverse domains [Feng et al. 2023, Biradar et al. 2024, Fawzy and Barsi 2025]. For instace, the work of [Feng et al. 2023]proposed for the automatic, rapid, and accurate detection of seepage in subway tunnels, a deep learning-based approach was proposed, resulting in the development of forty U-shaped semantic segmentation models. This work coupled the UNet and UNet++ architectures with six types of classification CNNs as encoders and utilized Grad-CAM++ for visual explanations. The UNet++ model with an EfficientNetB5 encoder demonstrated the best performance, achieving an 87.70% Intersection over Union (IoU). The study concluded that deepening and widening the encoder networks can boost model performance.

In a study aimed at accurately distinguishing between water and non-water pixels, several models, including Random Forest, SVM, FCN, and U-Net, were compared on a Sentinel-2 satellite dataset [Biradar et al. 2024]. The results demonstrated that U-net outperformed the compared models, achieving an accuracy of 86.93% and a mean IoU of 74.12% on the test set.

A CNN model employing the U-Net architecture was also developed for the semantic segmentation of VHR satellite imagery in an urban study area [Fawzy and Barsi 2025]. In this work, the model was trained to classify imagery into five urban classes (water, vegetation, bare soil, road, and building). When compared with the traditional Maximum Likelihood (ML) classification method, the U-Net approach proved superior, achieving an overall accuracy of 87.50%.

In addition to U-Net, recent studies have leveraged the DeepLabV3+ architecture to refine water body and shoreline extraction [Sun et al. 2024, Shen et al. 2025, Wang et al. 2026]. To address problems such as indistinct edges and the misclassification of shadows as water bodies, Sun et al. [Sun et al. 2024] proposed WaterDeep, a deep semantic segmentation network based on the DeepLabV3+ architecture. The approach uses a modified Xception network to extract low-level features and a densely connected

Atrous Spatial Pyramid Pooling (DASPP) module to aggregate multi-scale information. By fusing high- and low-level features in the decoder, the model demonstrated superior performance in distinguishing water bodies in complex urban environments, outperforming other architectures like UNet++ and SegNet and achieving an overall accuracy of 99.284%.

The study of [Shen et al. 2025] develops an optimized DeepLabV3+ framework for high-precision island shoreline extraction, using Koh Lan, Thailand, as a case study. By replacing the standard backbone with MobileNetV2 and integrating both a Strip Pooling layer and CBAM (Convolutional Block Attention Module) dual-attention mechanisms, the authors created a lightweight architecture capable of capturing long-range spatial dependencies and multi-scale contextual information. Validated on a self-built high-resolution drone dataset (1.5 cm spatial resolution), the model achieved a Pixel Accuracy (PA) of 98.7% and an Intersection over Union (IoU) of 96.2%, significantly outperforming traditional models like U-Net, FCN8, and SegNet. Furthermore, the optimized design reduced parameter counts by 88% and floating-point operations by 67%, establishing a highly efficient and cost-effective solution for real-time coastal monitoring on resource-constrained platforms such as UAVs.

Wang et al. [Wang et al. 2026] propose an enhanced DeepLabV3+ model specifically designed to improve the accuracy of wetland information extraction in the Liaohe River Estuary (LRE). To address the limitations of shallow architectures in capturing complex spatial structures and multi-scale feature characteristics, the study replaces the MobileNetV2 backbone with a ResNet-50 deep residual network. Additionally, the Atrous Spatial Pyramid Pooling (ASPP) module is optimized using a cross-fertilization mechanism in the parallel inflated convolutional branches, enabling more efficient multi-scale information fusion and improved identification of boundaries for irregular features like Suaeda salsa and Phragmites australis. Validated using Sentinel-2 imagery and field survey data, the proposed model achieved an overall accuracy of 97.6% and a Kappa coefficient of 0.971, outperforming traditional DeepLabV3+, ERFNET, and FCN architectures. These research results provide a scientific foundation and reliable data support for the sustainable management and long-term monitoring of the LRE wetland ecosystem.

The effectiveness of these models depends on both the available spectral information and their ability to generalize across different data sources. By benchmarking U-Net, SegFormer, and DeepLabV3+ architectures across seven coastal datasets, [Blais and Akhloufi 2025] demonstrated that spectral combinations beyond RGB, such as RGB-NIR and BR-NIR, provide a clearer distinction between land and sea. Each architecture, paired with VGG16, MiT-B5, and EfficientNet-B7 backbones, respectively, was trained and cross-tested across all datasets to evaluate how models transfer across varying resolutions, sensor types, and image qualities. The findings revealed a generalization asymmetry, where models trained on lower-resolution datasets generalized more successfully to high-resolution imagery than the reverse, emphasizing the necessity of cross-dataset evaluation to ensure real-world applicability.

Despite these advancements, the application of multispectral deep learning for hydrographic network segmentation under dense vegetation occlusion remains underexplored, as most studies focus on open water bodies or shorelines. Our work addresses this limitation by investigating how multispectral and index-based features influ-

ence the performance of U-Net and DeepLabV3+ models using deep backbones, specifically ResNet-152 and EfficientNet-B5. By contributing two separate datasets for the Itapemirim and Doce river basins and performing a rigorous cross-dataset analysis, this study validates the applicability of advanced deep learning frameworks for large-scale hydrological mapping in complex environmental conditions.

3. Materials and Methods

This section outlines the materials and methods employed in the study, including the datasets, backbone networks for feature extraction, performance metrics, and the loss functions used during the optimization process.

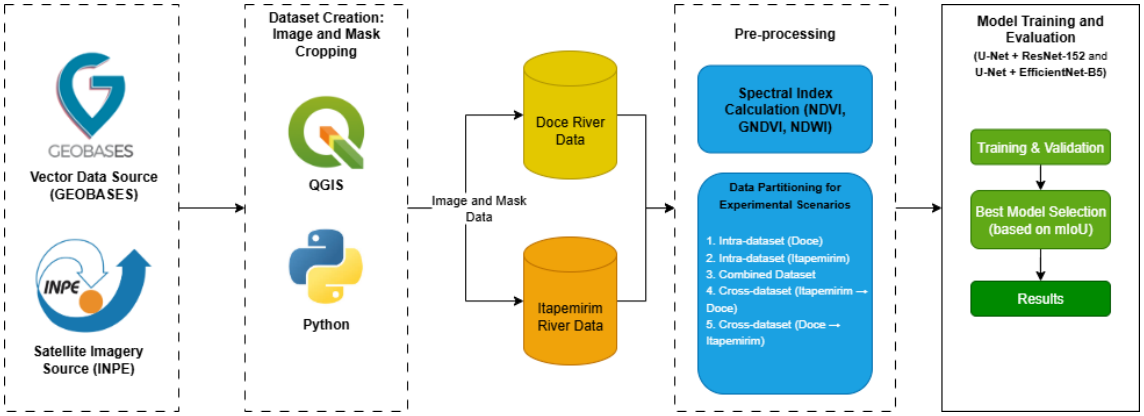


Figure 1. Overview of the proposed methodology for the comparative analysis of deep learning models for river network segmentation. The pipeline begins with the acquisition of vector data (GEOBASES) and satellite imagery (INPE), which are processed and cropped using QGIS and Python to generate two distinct datasets: Doce River and Itapemirim River. The pre-processing stage involves augmenting the image patches by calculating three spectral indices (NDVI, GNDVI, and NDWI), followed by partitioning the data for five distinct experimental scenarios. In the final stage, a comparative analysis is performed by training and validating U-Net and DeepLabv3+ architectures with different backbones with different backbones (ResNet-152 and EfficientNet-B5). The best model for each scenario is selected based on the mean Intersection over Union (mIoU) metric to produce the final results.

3.1. Datasets

The datasets for this study were created as part of this work by fusing multispectral satellite imagery with official geospatial vector data. The process involved two main types of data:

1. **Satellite Imagery:** Multispectral images from the CBERS-4 and CBERS-4A satellites, provided by the National Institute for Space Research (INPE), served as the base imagery.
2. **Official Hydrographic Reference Data:** High-precision vector maps from the Integrated Georeferenced Database of Espírito Santo (GEOBASES), produced by the Jones dos Santos Neves Institute (IJSN). These official cartographic products serve as the authoritative baseline for Land Use and Land Cover (LULC) in the region, providing validated classifications for water bodies, including reservoirs and main river channels.

The decision to utilize multispectral satellite imagery over Light Detection and Ranging (LiDAR) data was driven by operational feasibility and scalability. Although LiDAR represents the state-of-the-art for structural mapping under canopy due to its active laser penetration capabilities, its acquisition entails high operational costs and limited temporal frequency. In contrast, the CBERS constellation provides free, open-access data with high revisit rates, making it a sustainable and cost-effective solution for continuous large-scale monitoring of water resources in developing nations.

Preparing the final datasets for the model involved a multi-step process. We utilized the QGIS Geographic Information System [QGIS Development Team 2025] alongside custom Python scripts to fuse the input data and generate the core datasets. To enrich the deep learning model with robust multi-spectral information, we calculated and incorporated three spectral indices as supplementary input channels: the Normalized Difference Vegetation Index (NDVI), the Green Normalized Difference Vegetation Index (GNDVI), and the Normalized Difference Water Index (NDWI). The NDVI is a widely used index to quantify vegetation greenness and health. It is calculated using the near-infrared (NIR) and red bands [Rouse et al. 1974]:

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}. \tag{1}$$

The GNDVI is a modification of the NDVI that uses the green band instead of the red one, making it more sensitive to variations in chlorophyll concentration, which is useful for identifying vegetation near water bodies [Gitelson et al. 1996]:

$$\text{GNDVI} = \frac{(\text{NIR} - \text{Green})}{(\text{NIR} + \text{Green})}. \tag{2}$$

Finally, the NDWI is designed to highlight open water features while suppressing the signal from vegetation and soil. For this work, we used the formulation proposed by McFeeters [McFeeters 1996], which employs the green and NIR bands:

$$\text{NDWI} = \frac{(\text{Green} - \text{NIR})}{(\text{Green} + \text{NIR})}. \tag{3}$$

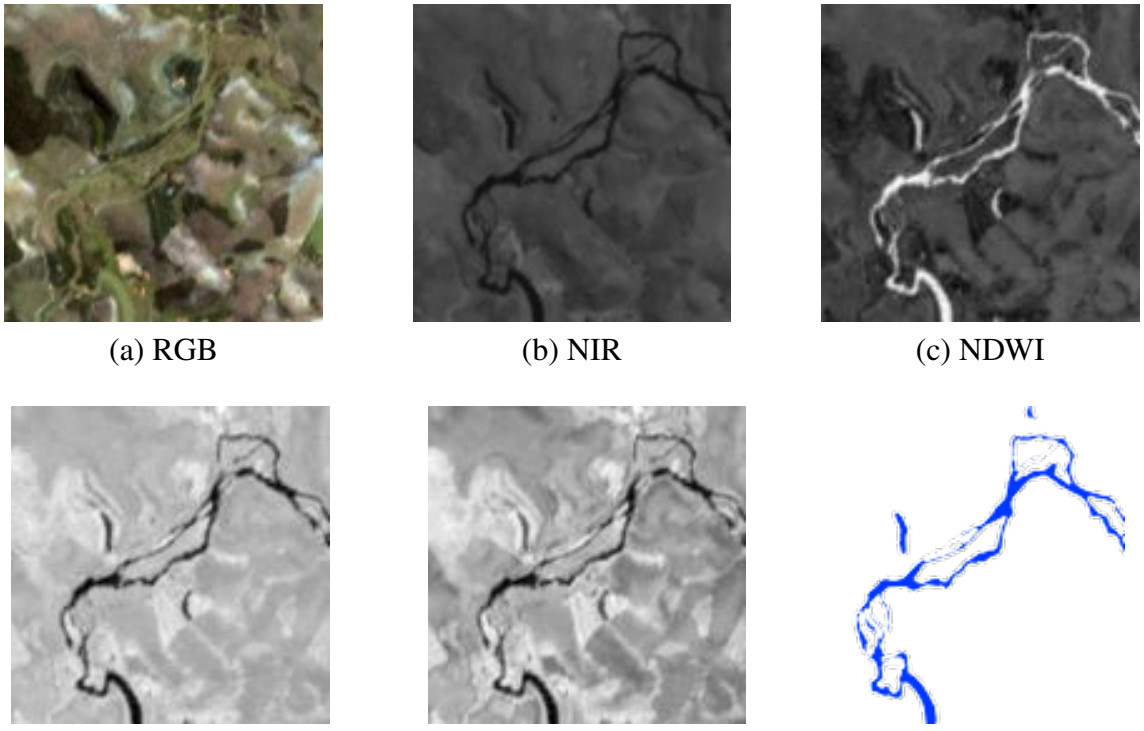
Thus, the final input for the network is a seven-channel image tensor: the first four channels are the RGB and Near-Infrared (NIR) bands from the satellite imagery, and the next three channels correspond to the NDVI, GNDVI, and NDWI indices. The ground truth labels were constructed through a rigorous rasterization pipeline. First, the official vector data were filtered to isolate features belonging strictly to the "water body" class. These vectors were then rasterized to match the exact spatial resolution (GSD) and alignment of the corresponding satellite imagery patches. This process resulted in binary segmentation masks where pixels representing water were labeled as 1 (foreground) and all other classes as 0 (background), ensuring spatial consistency between the spectral input and the target output.

The decision to incorporate spectral indices (NDVI, GNDVI, and NDWI) alongside raw RGB and NIR bands was based on the necessity of integrating domain-specific knowledge into the deep learning architecture. While raw bands provide fundamental

reflectance data, spectral indices act as a form of feature engineering that simplifies the identification of complex surface features. According to [Zhang et al. 2023], augmenting network inputs with multiple nonlinear spectral indices, specifically NDWI and GNDVI, boosts semantic segmentation performance and leads to superior contour localization when compared to standard three-channel (RGB) or four-channel (RGB-NIR) configurations. In the particular context of hydrological mapping, [Patil et al. 2024] demonstrated that integrating NDWI and NDVI improves the discrimination of water bodies from surrounding land features in complex agricultural and semi-arid regions.

Table 1. Summary of the datasets compiled for this study.

Dataset Name	Description	Image Size	Quantity
Itapemirim River	Dataset covering the Itapemirim River basin in the southern region of Espírito Santo.	128x128 pixels	140 images
Doce River	Dataset covering sections of the Doce River basin, a major river system in southeastern Brazil.	128x128 pixels	177 images
		Total	**317 images**



(a) RGB (b) NIR (c) NDWI

(d) GNDVI (e) NDVI (f) Ground Truth

Figure 2. Example of the multi-channel input data for the model. The first five images represent input features, while the last one is the target ground truth mask.

For this study, data were collected from two of the major rivers in the state of Espírito Santo: the Itapemirim River and the Doce River. A summary of these datasets is

presented in Table 1, with a few exemplars shown in Figure 2. The geographical locations of the Itapemirim and Doce River basins are presented in Figures 3 and 4, respectively.

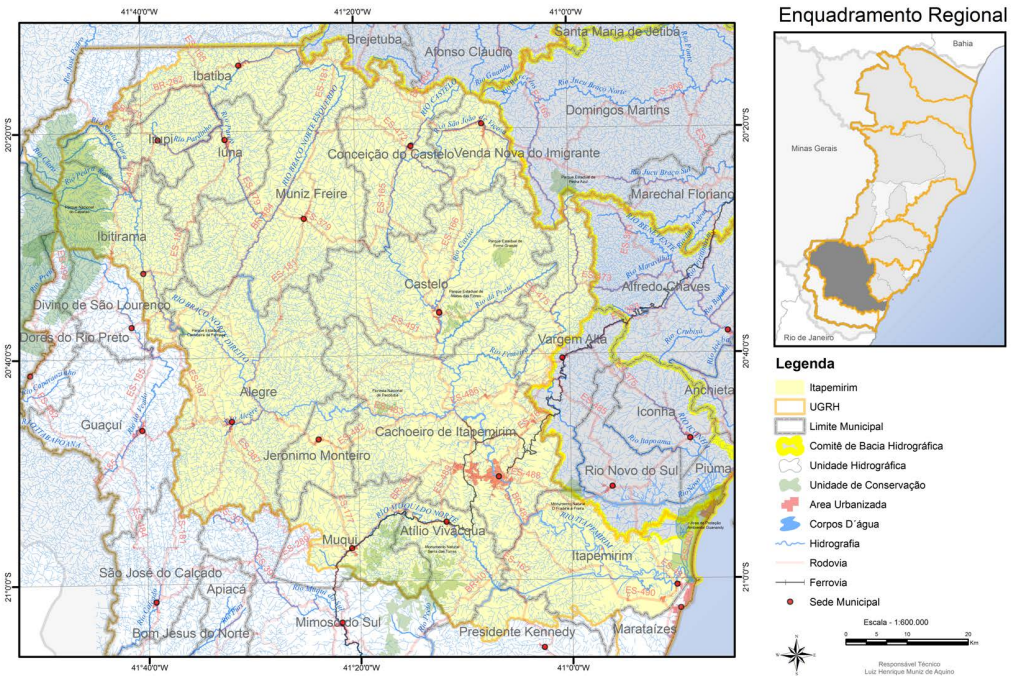


Figure 3. Itapemirim River Basin.

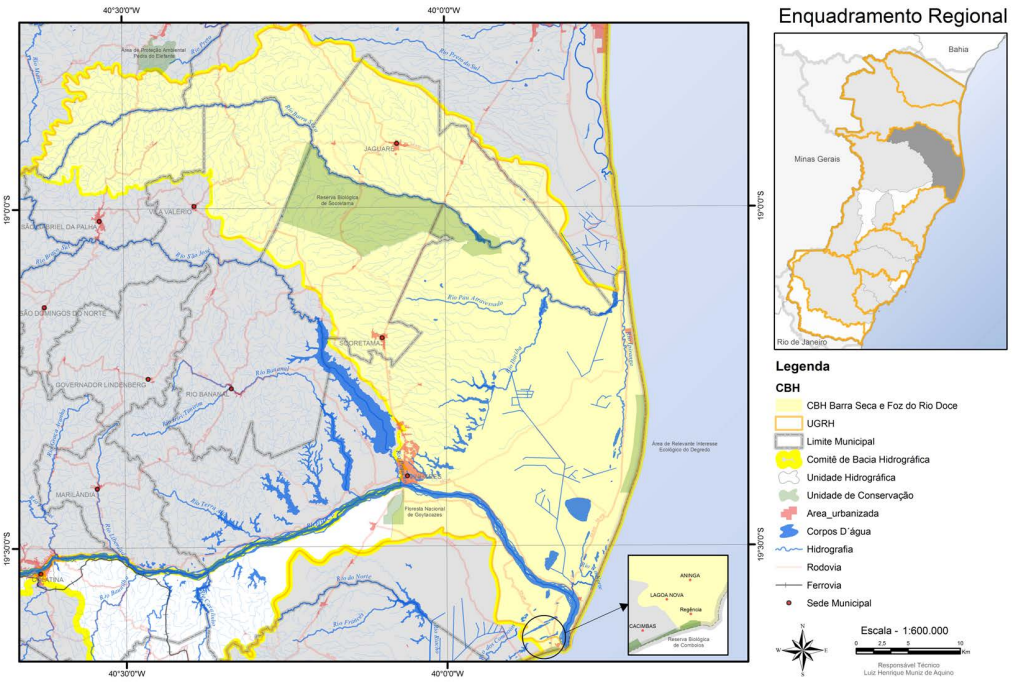


Figure 4. Doce River Basin.

3.2. Deep Learning Architectures

To address the challenge of river segmentation, we evaluate two state-of-the-art encoder-decoder architectures: U-Net and DeepLabv3+.

3.2.1. U-Net

Originally proposed for biomedical image segmentation, U-Net [Ronneberger et al. 2015] is characterized by its symmetric "U-shaped" architecture. It consists of a contracting path (encoder) that captures contextual information and an expansive path (decoder) that enables precise localization. A defining feature of U-Net is its skip connections, which directly concatenate high-resolution feature maps from the encoder to the decoder. This mechanism recovers spatial details lost during pooling operations, proving essential for delineating well-defined boundaries in hydrographic networks.

3.2.2. DeepLabv3+

DeepLabv3+ [Chen et al. 2018] combines the benefits of a spatial pyramid pooling module with an encoder-decoder structure. Its core component is the Atrous Spatial Pyramid Pooling (ASPP), which utilizes dilated (atrous) convolutions at multiple rates to capture multi-scale contextual information without significantly reducing spatial resolution. The decoder module then refines the segmentation results, sharpening object boundaries. This architecture is particularly effective at handling objects of varying scales and modeling long-range dependencies.

3.2.3. Pre-trained Backbones and Transfer Learning

To enhance feature extraction, we replace the standard encoders of both U-Net and DeepLabv3+ with powerful backbones pre-trained on the ImageNet dataset: ResNet-152 [He et al. 2016] and EfficientNet-B5 [Tan and Le 2019]. By employing these backbones, we leverage transfer learning [Yosinski et al. 2014], utilizing hierarchical features learned from millions of natural images. This strategy accelerates convergence and improves generalization, particularly in scenarios with limited labeled data like ours.

3.3. Metrics

To quantitatively evaluate the performance of the segmentation models, we adopted the Mean Intersection over Union (mIoU) metric, which is widely recognized as one of the most robust metrics in the literature [Everingham et al. 2010].

3.3.1. Mean Intersection over Union (mIoU)

The mIoU measures the similarity between the segmentation map predicted by the model and the ground truth mask. The basis of mIoU is the calculation of the Intersection over Union (IoU), also known as the Jaccard Index, for each class individually. IoU is the ratio

of the area of intersection to the area of union between the prediction and the ground truth. IoU is calculated at pixel level according to the equation:

$$\text{IoU} = \frac{\text{True Positives (TP)}}{\text{TP} + \text{False Positives (FP)} + \text{False Negatives (FN)}}. \tag{4}$$

In this context, True Positives (TP) represent pixels correctly identified as river, where both the prediction and the ground truth label indicate the presence of water. False Positives (FP) correspond to non-river pixels (background) incorrectly classified as river; here, the model predicts water, but the ground truth indicates background features such as vegetation, soil, or urban areas. Conversely, False Negatives (FN) refer to river pixels that the model failed to detect, where the ground truth indicates water but the model predicts background, often resulting from occlusion.

The IoU ranges from 0 (no overlap) to 1 (perfect overlap). To obtain the mIoU, the IoU is calculated for each of the classes in the dataset, and then the mean of these values is calculated. In our binary segmentation problem, the classes are "water body" and "background". The mIoU will, therefore, be the average of the IoU scores obtained for these two classes. This metric is particularly effective because it penalizes both under-segmentation (high FN) and over-segmentation (high FP), providing a balanced assessment of the model's quality.

3.4. Loss Function

The choice of the loss function is critical for guiding the model's optimization toward accurate segmentation, especially given the class imbalance inherent in mapping thin river networks (water vs. background). To leverage the strengths of different metrics, we employed a combined loss function integrating two standard components: Binary Cross-Entropy (BCE) Loss, which promotes overall pixel-wise classification accuracy, and Dice Loss, which is robust against class imbalance by optimizing for spatial overlap.

Binary Cross-Entropy (BCE) measures the dissimilarity between two probability distributions: the model's predicted output (the probability of each pixel belonging to the "water" class) and the true distribution (the ground truth, where each pixel is either 0 or 1). BCE heavily penalizes confident but incorrect predictions, optimizing for accurate pixel-wise classification. The formula is given by:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)], \tag{5}$$

where N is the total number of pixels, y_i is the true label (0 or 1) of pixel i, and $\hat{y}_i$ is the model's predicted probability for pixel i.

Dice Loss is derived from the Dice Coefficient, which measures the overlap between two samples. For segmentation, it has become extremely popular for effectively handling class imbalance (e.g., few water pixels compared to the background) [Milletari et al. 2016]. Instead of evaluating each pixel individually like BCE, Dice Loss directly optimizes the overlap between the predicted and ground truth masks, focusing on the spatial consistency of the segmentation. Formally, the Dice Loss is defined as:

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i + \epsilon}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon}, \tag{6}$$

where y_i and $\hat{y}_i$ are the same as in BCE, and ϵ is a smoothing term to prevent division by zero.

The final loss function used was the simple average of BCE and Dice Loss, given by

$$L_{Combined} = \frac{1}{2}(L_{BCE} + L_{Dice}). \tag{7}$$

This hybrid approach combines the benefits of both: the stability and good pixel-level convergence of BCE with the robustness of Dice Loss regarding class imbalance and spatial structure.

4. Results and Discussion

To validate our approach and evaluate the model's performance and generalization capabilities, we designed five main training and testing scenarios:

1. Intra-dataset (Itapemirim River): The model was trained and evaluated using only the Itapemirim River dataset.
2. Intra-dataset (Doce River): The same procedure was performed using only the Doce River dataset.
3. Combined Dataset: The model was trained using the merged Itapemirim and Doce river datasets to assess performance on a more diverse data pool.
4. Cross-dataset Generalization (Itapemirim $\rightarrow$ Doce): To test the ability to generalize to a distinct geographical area, the model was trained exclusively on the Itapemirim River dataset and evaluated on the entire Doce River dataset.
5. Cross-dataset Generalization (Doce $\rightarrow$ Itapemirim): Inverting the previous scenario, the model was trained on the Doce River dataset and evaluated on the entire Itapemirim River dataset.

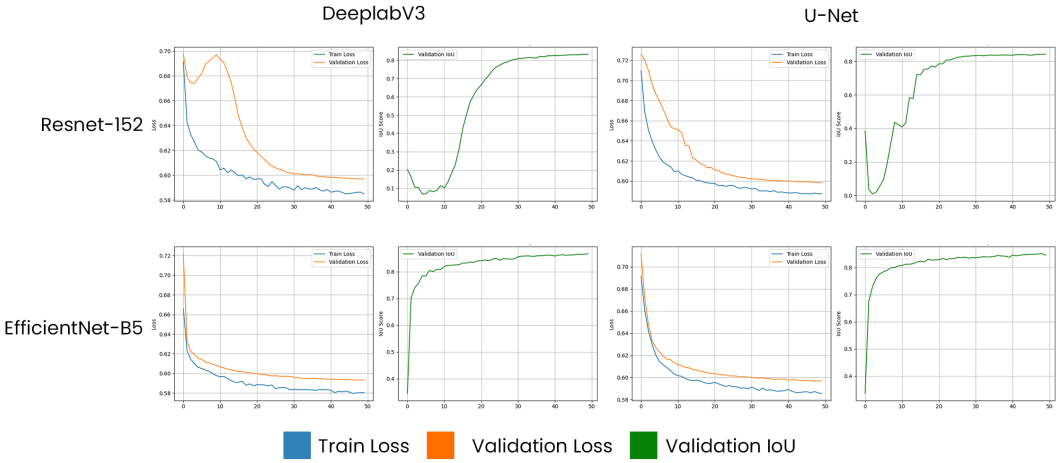
As in [Weng et al. 2023], Scenarios 1–3 follow an 80/20 train–test split, whereas in the cross-dataset experiments, the ratio between the training and testing sets is intrinsically defined by the size of Itapemirim and Doce River datasets. The proposed method was evaluated on each scenario for the U-Net and DeepLabv3 architectures for comparison purposes. Both were trained using two different pre-trained backbones: EfficientNet-B5 and ResNet-152. To ensure statistical reliability and account for stochastic variations in initialization, each experimental configuration was repeated 10 times. Consequently, the quantitative results reported herein represent the mean and standard deviation of these runs.

All models used a seven-channel input configuration consisting of the three RGB channels plus NIR, NDVI, GNDVI, and NDWI, as discussed in Section 3.1, and were implemented using the PyTorch library[1]. The main hyperparameters used for training are detailed in Table 2. The loss function adopted was a combination of Binary Cross-Entropy and Dice Loss, as explained in Section 3.4.

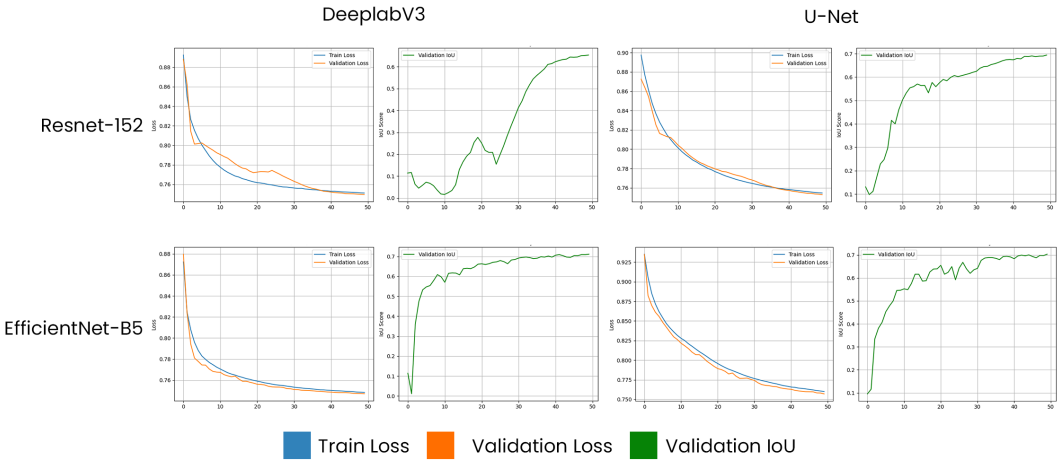
[1]To ensure reproducibility, the source code, trained weights, and dataset are available at: `https://github.com/duvrdx/dnn-segmentation-rivers`.

Table 2. Hyperparameters used for model training.

Hyperparameter	Value
Optimizer	Adam
Learning Rate	1×10^{-4}
Encoder Depth	5
Encoder Weights	imagenet
Activation	Sigmoid
Epochs	50
Batch Size	8
Loss Function	BCE + Dice Loss
Input Size	128x128 pixels
Input Channels	7 (RGB, NIR, NDVI, GNDVI, NDWI)

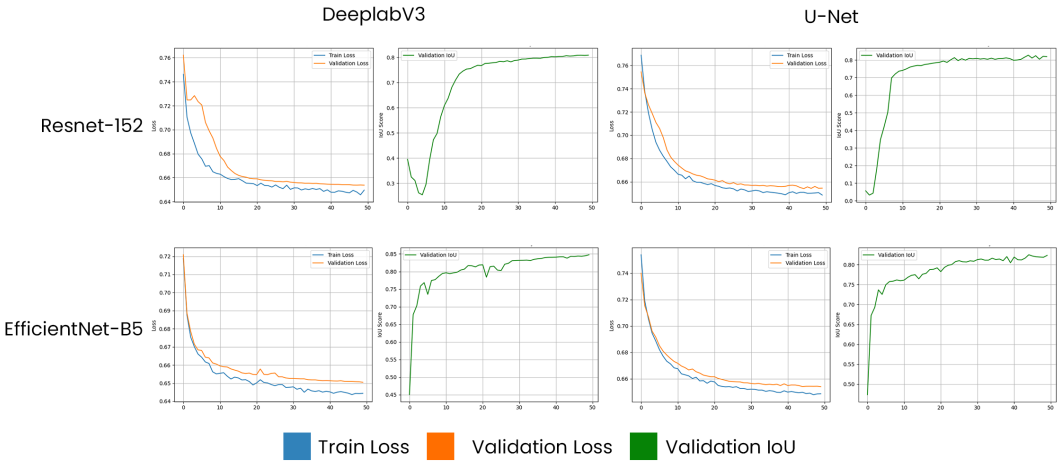


(a) Intra-dataset (Doce River)

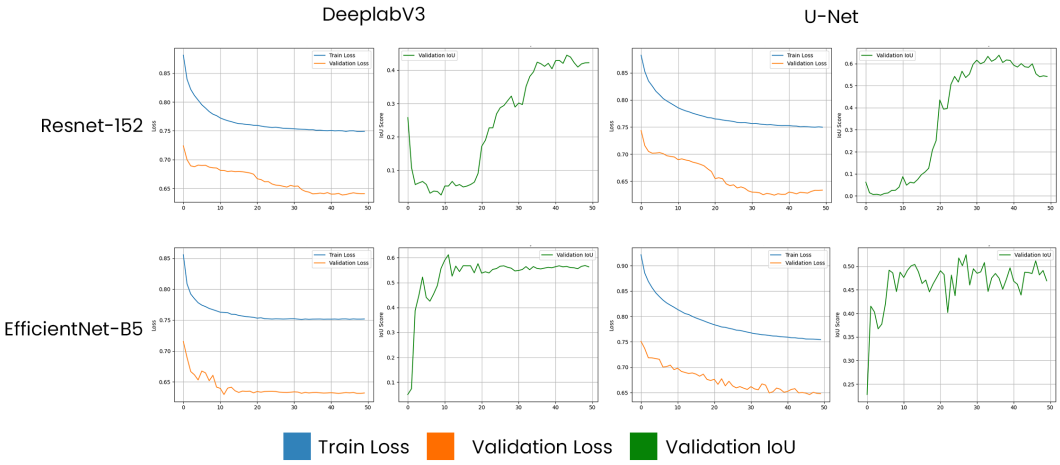


(b) Intra-dataset (Itapemirim River)

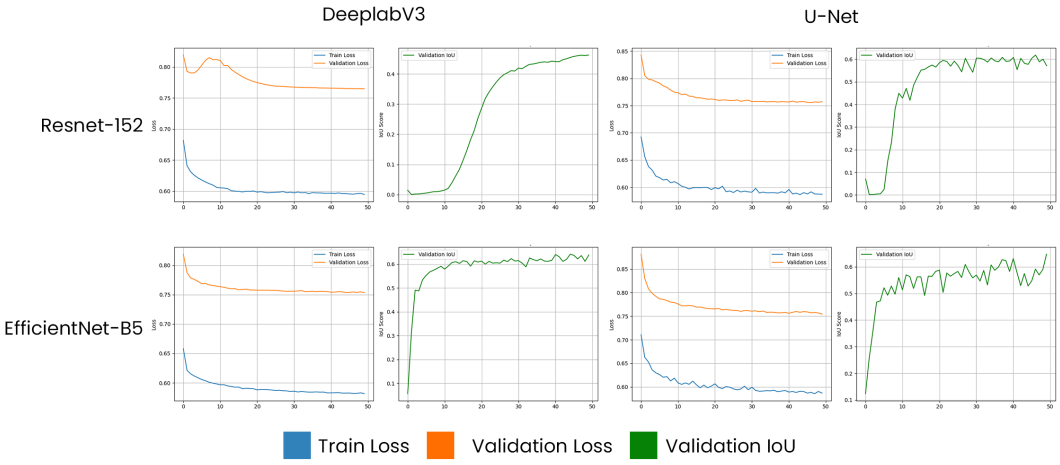
Figure 5. Intra-dataset training results using EfficientNet-B5 and ResNet152 as the backbone. In (a) training and validation on the Doce River dataset, and in (b) on the Itapemirim River dataset.



(a) Combined Dataset



(b) Cross-dataset Generalization (Itapemirim → Doce)



(c) Cross-dataset Generalization (Doce → Itapemirim)

Figure 6. Results for combined training and cross-dataset generalization. In (a) the result for the combined dataset, in (b) the generalization from Itapemirim to Doce, and in (c) the generalization from Doce to Itapemirim.

4.1. Training Curves

Figure 5 illustrates the training and validation dynamics for intra-dataset scenario, the Doce River (Figure 5a) and Itapemirim River (Figure 5b) datasets. The results are organized by backbone architecture in rows, while the columns distinguish the performance of the two model architectures: the two leftmost columns present the loss function and mIoU scores for DeepLabv3+, and the two rightmost columns show the corresponding metrics for U-Net. Each plot demonstrates a standard convergence behavior where the loss function minimizes as the mIoU score increases over 50 epochs. Notably, the ResNet-152 backbone exhibited initial stochastic instability when paired with both architectures, characterized by early-stage fluctuations before reaching a stable convergence. This behavior is consistent with the challenges of using deeper residual structures to capture intricate features in high-resolution hydrological segmentation.

Figure 6 illustrates the training and validation dynamics for the combined and cross-dataset experimental scenarios. The results for the combined-dataset (Figure 6a), which integrates samples from both river basins, show a standard convergence profile with the loss function gradually minimizing as the mIoU scores increase. Consistent with the intra-dataset results, the ResNet-152 backbone displayed early-stage stochastic instability before reaching a stable state.

In contrast, the cross-dataset generalization tests evaluating the Itapemirim-trained model on the Doce River basin (Figure 6b) and the Doce-trained model on the Itapemirim basin (Figure 6c) reveal a significant performance gap. In these scenarios, the mIoU stabilized at approximately 63%, representing a substantial decrease compared to the intra-dataset benchmarks. The most anomalous result occurred in the Itapemirim $\rightarrow$ Doce generalization test (Figure 6(b)), where the mIoU score started high but then deteriorated over time, even as the training loss decreased. This divergence between the decreasing training loss and the deteriorating cross-dataset mIoU is a clear indicator of domain shift, a phenomenon extensively discussed in domain adaptation literature [Ganin et al. 2016]. As the model continues to minimize the loss function on the source domain (Itapemirim), it begins to overfit to source-specific spatial-spectral features, such as particular riparian vegetation textures or water turbidity levels, that are not representative of the target domain (Doce). Consequently, learned feature representations become less transferable, leading to performance degradation on the unseen target domain. This confirms that minimizing source risk does not guarantee target generalization when significant distributional shifts exist [Ganin et al. 2016].

4.2. Quantitative Results

Table 3 organizes the quantitative performance results, with experimental scenarios presented in rows and model-backbone combinations in columns. The primary columns partition the DeepLabv3+ (right) and U-Net (left) architectures, each subdivided into ResNet-152 and EfficientNet-B5 backbones. The rows detail the intra-dataset evaluations for the Doce and Itapemirim rivers, the combined dataset scenario, and the cross-dataset generalization tests. Values represent the mean mIoU scores ± standard deviation over ten runs.

In the intra-dataset scenarios, DeepLabv3+ paired with the ResNet-152 backbone achieved superior performance, reaching 86.73% for the Doce River and 71.49% for the

Itapemirim River. For the U-Net architecture, the EfficientNet-B5 backbone consistently outperformed ResNet-152, achieving a peak score of 84.86% on the Doce River dataset.

The combined dataset scenario demonstrated robust and stable performance, with the DeepLabv3 (ResNet-152) model achieving a top result of 84.29% mIoU. This indicates that the models effectively integrated features from geographically distinct regions without a significant loss in accuracy compared to the individual intra-dataset benchmarks.

In the cross-dataset generalization tests, the performance of the most robust configurations, specifically DeepLabv3 with ResNet-152 and U-Net with EfficientNet-B5, reached a comparable plateau, with mIoU scores for both configurations gravitating around 63%. This statistical parity suggests that, despite their distinct architectural designs, both models achieved a similar level of generalization capability when transferred between the Itapemirim and Doce river basins.

A potential concern regarding the use of deep backbones such as ResNet-152 and EfficientNet-B5 on a dataset of 317 images is the risk of overfitting. However, this risk was mitigated through two primary strategies. First, we employed transfer learning with models pre-trained on the ImageNet dataset [Fei-Fei et al. 2009]. Transfer learning is a critical technique for applying deep learning models to domains with limited training data, as it allows the network to leverage robust feature representations learned from large-scale datasets, thereby preventing the model from merely 'memorizing' a small sample [Yap et al. 2023]. Second, the U-Net architecture was specifically designed to yield high-quality segmentations even with limited training data [Bardis et al. 2020, Tiwari and Saraswat 2024].

The cross-dataset generalization tests provide a critical assessment of the model's robustness. The significant performance drop of approximately 20 percentage points compared to the combined dataset results highlights the challenges of domain shift. This outcome indicates that, while the models generalize well intra-domain, they are not robust to inter-domain variations given the current data scale. Consequently, the results suggest a continued dependence on including site-specific training samples to achieve high precision in distinct geographical regions.

Table 3. Comparison of Mean Intersection over Union (mIoU) scores (%) on the test set for U-Net and DeepLabv3 models. Results are reported as *Mean* ± *Standard Deviation* over 10 runs. The best result is shown in bold.

Experimental Scenario	U-Net		DeepLabv3	
	ResNet-152	EfficientNet-B5	ResNet-152	EfficientNet-B5
Intra-dataset (Doce River)	84.34 ± 0.44	84.86 ± 0.39	**86.73 ± 0.21**	80.97 ± 2.80
Intra-dataset (Itapemirim River)	69.68 ± 0.59	70.94 ± 0.74	**71.49 ± 0.30**	65.31 ± 0.20
Combined Dataset	81.90 ± 0.48	82.25 ± 0.51	**84.29 ± 0.27**	80.90 ± 0.22
Cross-Dataset Generalization Tests				
Itapemirim → Doce	59.15 ± 3.83	**63.94 ± 3.81**	**63.47 ± 2.89**	47.02 ± 2.74
Doce → Itapemirim	62.78 ± 1.98	**63.03 ± 1.19**	**63.96 ± 0.88**	50.02 ± 4.11

4.3. Qualitative Analysis

Regarding the architectural comparison, while DeepLabv3+ achieved slightly higher scores in some scenarios, a visual inspection revealed that its segmentation masks tended to over-smooth complex boundaries. In contrast, the U-Net architecture demonstrated superior capability in preserving fine-grained details of narrow tributaries, which is critical for the topological correctness of hydrographic networks, even if it resulted in a marginally lower mIoU due to boundary noise.

Figure 7 provide a qualitative assessment of the segmentation results from the models validation phase. For the intra-dataset and combined dataset scenarios, the visual predictions corroborate the quantitative data, demonstrating that the model adapts well to the complexities of the hydrological networks. While some minor omissions are present, the rate of false positives is low.

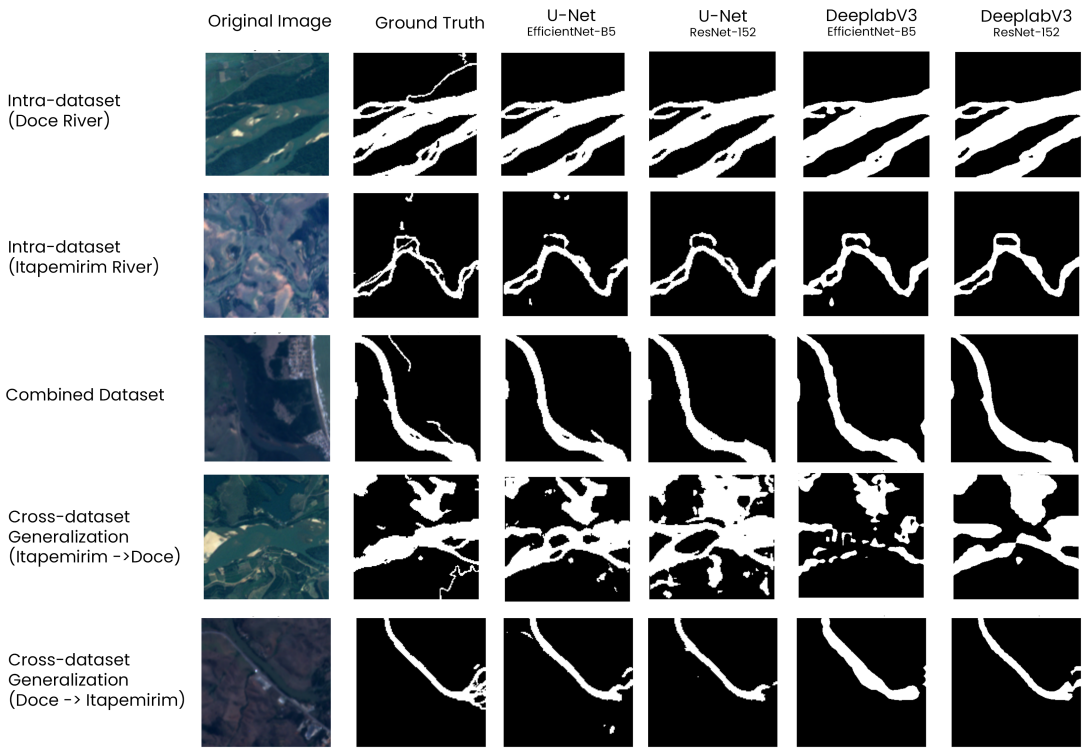


Figure 7. Sample segmentation results for U-Net and DeepLabv3 models with EfficientNet-B5 and ResNet-152 backbones.

The cross-dataset generalization tests, however, reveal an interesting asymmetry. In the Itapemirim $\rightarrow$ Doce scenario, the model trained on the more complex hydrological network (Itapemirim) generalizes effectively to the simpler one (Doce). As depicted in the example, it correctly segments the river and adjacent water bodies. The inverse scenario does not yield the same success. When trained on the simpler Doce River dataset, the model accurately delineates the main river channel of the Itapemirim River but struggles to segment its more intricate and branching areas, even while maintaining a low false positive rate.

4.4. Implications for Decision Support Systems

The integration of accurate river segmentation models into Decision Support Systems (DSS) represents a significant advancement for flood risk management and hydrological modeling. While quantitative forecasting models, such as Multilayer Perceptrons optimized by Neural Architecture Search, have demonstrated high precision in predicting river flow rates (m^3/s) [Souza et al. 2025], they typically lack explicit spatial modeling. The U-Net architecture proposed in this work complements such quantitative predictions by providing precise, real-time extraction of channel geometries and flood masks.

By ingesting the binary masks generated by our model into Geographic Information Systems (GIS), public agencies can automate the calibration of hydraulic models (e.g., HEC-RAS) that rely on updated roughness coefficients and channel boundaries. Furthermore, in flood warning scenarios, the synergy between flow forecasting [Souza et al. 2025] and automated segmentation enables the rapid generation of flood inundation maps, allowing civil defense authorities to identify vulnerable urban areas with greater agility than traditional manual vectorization methods.

Moreover, the precise delineation of river boundaries directly supports environmental compliance and land use planning, specifically regarding Permanent Preservation Areas (APPs). According to the Brazilian Forest Code, the width of the protected riparian buffer is determined by the width of the watercourse itself. By automating the extraction of river banks, the proposed model enables the dynamic and accurate generation of these buffer zones in GIS. This capability facilitates the large-scale monitoring of illegal occupation or deforestation within protected areas, providing environmental agencies with a robust tool to enforce legislation and promote ecosystem conservation.

5. Conclusion

This paper presented a comparative evaluation of U-Net and DeepLabv3+ architectures with EfficientNet-B5 and ResNet-152 backbones for the semantic segmentation of river networks under vegetation occlusion. Using datasets derived from georeferenced data of the Doce and Itapemirim rivers, the models were trained and tested across multiple scenarios, including intra-dataset, combined, and cross-dataset experiments.

The results demonstrated the effectiveness of transfer learning for hydrological segmentation, revealing distinct strengths for each architecture. While DeepLabv3+ with ResNet-152 achieved the highest mIoU in all dataset scenario. These findings indicate that while DeepLabv3+ offers robust generalization, and U-Net remains a powerful tool for capturing fine-grained details. Although cross-dataset tests showed lower performance (mean mIoU around 63%), both models retained their ability to delineate river structures across distinct regions, supporting their practical applicability in large-scale hydrological mapping.

The integration of multispectral imagery with derived indices (NDVI, GNDVI, and NDWI) proved essential for detecting water bodies beneath vegetation cover. This confirms that extending the spectral range beyond RGB channels enhances feature discrimination and the accuracy of hydrographic segmentation. The combination of transfer learning, pre-trained backbones, and the encoder–decoder structures enabled precise boundary reconstruction even in complex river networks. From a theoretical standpoint,

this work applied the Task–Technology Fit framework to assess the suitability of deep learning architectures for automated hydrographic network extraction. The results validate the effectiveness of transfer learning in this context, indicating potential for broader application in environmental monitoring and geospatial analysis.

The proposed method offers a reproducible and efficient approach to automate drainage network extraction in remote sensing data, contributing to hydrological modeling, flood risk assessment, water resource management, and ecosystem conservation. Future research should focus on expanding the dataset to other biomes and climatic conditions, exploring newer segmentation models such as Vision Transformers or Siamese networks, and integrating multimodal data (e.g., LiDAR and SAR) to enhance robustness in occluded environments. It is also necessary to include a quantitative analysis of variance across multiple runs, reporting mean and standard deviation of mIoU values for different random seeds, and to apply statistical significance tests (e.g., t-tests) when comparing model backbones. Such procedures would strengthen the statistical rigor and reinforce the reliability of comparative results. Additional directions include developing real-time processing and federated learning strategies to support distributed hydrological mapping.

Acknowledgements

The authors thank IFES and FAPES for grant No. 1048/2025 (P: 2025-CGNQ0), project "DI 016/2025 - IntegraCAR: Integração do Cadastro Ambiental Rural no Estado do Espírito Santo". Professor Komati thanks CNPq for the DT-2 grant (302726/2023-3) and grant 407742/2022-0, as well as FAPES for project 1023/2022 (P:2022-8TZV6). The authors acknowledge the use of Gemini to assist with language translation and grammar revision. The authors take full responsibility for the content and integrity of this article.

References

Bardis, M., Houshyar, R., Chantaduly, C., Ushinsky, A., Glavis-Bloom, J., Shaver, M., Chow, D., Uchio, E., and Chang, P. (2020). Deep learning with limited data: organ segmentation performance by U-Net. *Electronics*, 9(8):1199.

Biradar, R. L., Thatipalli, S., Mucharla, A., Adepu, S., and Mandava, P. (2024). Detection of water bodies using satellite imagery based on deep learning. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 12(5).

Blais, M.-A. and Akhloufi, M. A. (2025). Benchmarking coastal boundary datasets in deep learning applications. *Earth Science Informatics*, 18(4):520.

Cao, H., Tian, Y., Liu, Y., and Wang, R. (2024). Water body extraction from high spatial resolution remote sensing images based on enhanced U-Net and multi-scale information fusion. *Scientific Reports*, 14(1):16132.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.

Everingham, M., Eslami, S. M. A., Gool, L. V., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338.

Fawzy, M. and Barsi, A. (2025). A U-Net model for urban land cover classification using vhr satellite images. *Periodica Polytechnica Civil Engineering*, 69(1):98–108.

Fei-Fei, L., Deng, J., and Li, K. (2009). Imagenet: Constructing a large-scale image database. *Journal of vision*, 9(8):1037–1037.

Feng, S. J., Feng, Y., Zhang, X. L., and Chen, Y. H. (2023). Deep learning with visual explanations for leakage defect segmentation of metro shield tunnel. *Tunnelling and Underground Space Technology*, 136:105107.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. (2016). Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030.

Gitelson, A. A., Kaufman, Y. J., and Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote sensing of Environment*, 58(3):289–298.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

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.

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-Net: fully convolutional neural networks for volumetric medical image segmentation. *arXiv preprint arXiv:1606.04797*, pages 565–571.

Osias, A. C. F., Schaefer, M. A. R., Veloso, G. V., de Oliveira, H. N., and Reis, J. C. S. (2024). Interpretable approaches for land use and land cover classification. In *Proceedings of Simpósio Brasileiro de Sistemas de Informação (SBSI'24)*, SBSI'24, New York, NY, USA. Association for Computing Machinery.

Patil, P. P., Jagtap, M. P., Khatri, N., Madan, H., Vadduri, A. A., and Patodia, T. (2024). Exploration and advancement of NDDI leveraging NDVI and NDWI in Indian semi-arid regions: A remote sensing-based study. *Case Studies in Chemical and Environmental Engineering*, 9:100573.

QGIS Development Team (2025). *QGIS Geographic Information System*. Open Source Geospatial Foundation. Version 3.36.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer.

Rouse, J. W., Haas, R. H., Schell, J. A., and Deering, D. W. (1974). Monitoring vegetation systems in the great plains with erts. *NASA special publication*, 351:309.

Rusnák, M., Goga, T., Michaleje, L., Šulc Michalková, M., Máčka, Z., Bertalan, L., and Kidová, A. (2022). Remote sensing of riparian ecosystems. *Remote Sensing*, 14(11):2645.

Shen, J., Guo, Z., Zhang, Z., Plathong, S., Jantharakhantee, C., Ma, J., Ning, H., and Qi, Y. (2025). Remote sensing shoreline extraction method based on an optimized

deeplabv3+ model: A case study of koh lan island, thailand. *Journal of Marine Science and Engineering*, 13(4):665.

Souza, E. H. P., Oliveira, V. M. d., Andrade, J. O., and Komati, K. S. (2025). Previsão de vazão de rios usando rede perceptron multi-camada otimizada por neural architecture search. *Tecnia: Revista de Educação, Ciência e Tecnologia do IFG*, 10(Edição Especial 1).

Sun, D., Gao, G., Huang, L., Liu, Y., and Liu, D. (2024). Extraction of water bodies from high-resolution remote sensing imagery based on a deep semantic segmentation network. *Scientific Reports*, 14(1):14604.

Tan, M. and Le, Q. V. (2019). EfficientNet: rethinking model scaling for convolutional neural networks. In Chaudhuri, K. and Salakhutdinov, R., editors, *Proceedings of the 36th International Conference on Machine Learning (ICML)*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR.

Tiwari, T. and Saraswat, M. (2024). Analysis of UNet-Based semantic segmentation models. In *International Conference on Computing and Machine Learning*, pages 421–431. Springer.

Viana, A. B., da Silva, R. A., and de Oliveira, V. d. P. S. (2024). Uso racional de água de reuso ou potável na indústria. *Boletim do Observatório Ambiental Alberto Ribeiro Lamego*, 18(2):17–35.

Wang, Y., He, J., Wang, C., and Zhang, W. (2026). Wetland information extraction method based on improved deeplabv3+ in liaohe river estuary. *Wetlands*, 46(1):11.

Weng, L., Pang, K., Xia, M., Lin, H., Qian, M., and Zhu, C. (2023). Sgformer: A local and global features coupling network for semantic segmentation of land cover. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:6812–6824.

Yap, H. Y., Choo, Y.-H., Mohd Yusoh, Z. I., and Khoh, W. H. (2023). An evaluation of transfer learning models in EEG-based authentication. *Brain informatics*, 10(1):19.

Yosinski, J., Clune, J., and O. L. Sinsap, Yoshua Bengio, H. L. (2014). How transferable are features in deep neural networks? In *Advances in neural information processing systems*, volume 27.

Yuan, K., Zhuang, X., Schaefer, G., Feng, J., Guan, L., and Fang, H. (2021). Deep-learning-based multispectral satellite image segmentation for water body detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:7422–7434.

Zhang, Y., Yang, R., Dai, Q., Zhao, Y., Xu, W., Wang, J., and Wang, L. (2023). Boosting semantic segmentation of remote sensing images by introducing edge extraction network and spectral indices. *Remote Sensing*, 15(21):5148.