Pernambuco Waterbody Dataset (PWD): A High-Resolution Multi-Source Dataset for Deep Learning-Based Waterbody Segmentation in Tropical and Semi-Arid Regions

Tiago Prudencio Silvano [1], Reginaldo Pereira da Silva Filho [2], Joel Passos [3], Simone Sayuri Sato [4], Cleber Zanchettin [5]

[1] Dept. of Cartographic Engineering, Federal University of Pernambuco (UFPE), Recife – PE, Brazil – tiago.prudencio@ufpe.br
[2] Brazilian Army Geographic Service (DSG), Brasília – DF, Brazil – reginaldosilva.filho@eb.mil.br
[3] Dept. of Cartographic Engineering, Federal University of Pernambuco (UFPE), Recife – PE, Brazil – joel.passos@ufpe.br
[4] Dept. of Cartographic Engineering, Federal University of Pernambuco (UFPE), Recife – PE, Brazil – simone.sato@ufpe.br
[5] Informatics Center, Federal University of Pernambuco (UFPE), Recife – PE, Brazil – cz@cin.ufpe.br

Keywords: Remote Sensing Images, Water Body Extraction, Deep Learning, Dataset Construction.

Abstract

Accurate extraction of water bodies from remote sensing imagery is essential for environmental monitoring, water resource management, and hydrological applications. However, the performance of deep learning models for water segmentation depends on the availability of representative datasets that capture diverse environmental and spectral conditions, particularly in tropical and semi-arid regions that remain underrepresented in existing datasets. This study presents the Pernambuco Waterbody Dataset (PWD), a multi-source dataset comprising aerial and satellite remote sensing imagery for water body segmentation. The dataset covers the state of Pernambuco, Brazil, including tropical and semi-arid environments associated with the Atlantic Forest and Caatinga biomes. The dataset includes high-resolution aerial imagery (0.5 m) from the Pernambuco Tridimensional Program (PE3D) and Sentinel-2A imagery (10 m), with manually annotated water bodies generated by cartographic specialists. The dataset was constructed through data acquisition, preprocessing, manual annotation, mask generation, patch extraction (512 × 512 pixels), and division into training, validation, and test subsets. The first version includes 51,743 aerial patches and 15,321 Sentinel-2A patches. To validate the dataset, U-Net, U-Net++, and DeepLabV3+ architectures with ResNet and EfficientNet backbones were evaluated using Recall, Precision, F1-score, and IoU metrics. The best performance was achieved by U-Net++ (ResNet34) for aerial imagery (IoU 0.946) and U-Net (ResNet34) for Sentinel-2A imagery (IoU 0.871). Overall, the proposed dataset provides a robust benchmark for advancing deep learning-based water body extraction using multi-source remote sensing data.

1. Introduction

Water plays a vital role in human life and has significant ecological importance (Attya et al., 2025; Chen et al., 2024; Jonnala et al., 2025). Water resources directly influence human health (Knight et al., 2022; Yu et al., 2021), impacting quality of life, and constitute essential sources for agricultural production, industrial activities, and energy generation (Lei et al., 2025; Nagaraj and Kumar, 2023).

Surface water resources cover approximately 3% of the global land surface (Klein et al., 2017). In the Brazilian context, it is estimated that around 240,000 water bodies have been mapped, covering an area of 173,749.56 km², which corresponds to approximately 2% of the national territory (National Water and Sanitation Agency, 2022).

The investigation, detection, and mapping of these areas are crucial for water resource management, flood risk assessment, coastal zone planning, environmental studies, and land use planning. In this context, water body extraction from remote sensing imagery has become one of the main approaches for obtaining information on the spatial distribution of surface water bodies (Zhang et al., 2023).

Over the past decade, researchers have proposed approaches ranging from traditional computer vision methods, such as thresholding techniques applied to spectral water indices (Du et al., 2016; Wu et al., 2018; Zhai et al., 2015), to Machine Learning (ML) methods (Bangira et al., 2019; Hibjur Rahaman et al., 2023; Li et al., 2021).

Deep Learning (DL) methods have demonstrated superior performance in various computer vision tasks, including classification, object detection, and semantic segmentation (Yuan et al., 2021). Since water body extraction in remote sensing imagery is a typical image segmentation task, deep learning has been widely recognized as an effective approach for extracting and interpreting spatial and semantic features, mainly through Deep Neural Networks (DNNs) (Li et al., 2025, 2022; Xiang et al., 2025).

Semantic segmentation networks are capable of learning texture, spectral, and other high-dimensional features from labeled datasets and subsequently using these features to extract regions of interest in images (Xiang et al., 2025). These methods eliminate the need for manual threshold tuning, scale efficiently to large datasets, and exhibit greater flexibility and generalization capability, which explains their widespread adoption in water body extraction studies (Chen et al., 2024).

Despite the strong performance of DNN models, the main bottleneck is the lack of ground truth labels for model training and testing. DNN models require large volumes of training images with corresponding high-quality labels, whereas labeling pixel-wise ground truth masks is a laborious and time-consuming task (Vongkusolkit et al., 2023). Similarly, Portalés-Julià et al. (2023) highlight that the training process of deep learning methods requires a huge amount of labeled data, namely annotated reference masks that are fed into the network to learn the relevant features and the best mapping to classify water.

However, many publicly available datasets present limitations related to geographic diversity, spatial resolution, and environmental variability. In particular, tropical and semi-arid environments exhibit significant spectral and morphological variability of water bodies, which makes the segmentation task more challenging. In these regions, the presence of seasonal vegetation, intermittent water bodies, variations in turbidity, and changes in illumination conditions over time contribute to increasing the complexity of identifying these features in remote sensing imagery.

Furthermore, there is a scarcity of datasets specifically focused on the Brazilian territory, especially in regions characterized by high environmental heterogeneity, such as the Northeast of Brazil. This limitation hinders not only the development but also the systematic evaluation of water body segmentation methods in complex, representative environmental scenarios.

In this context, this study makes four main contributions:

1. The introduction of the Pernambuco Waterbody Dataset (PWD), a high-resolution multi-source remote sensing dataset for water body extraction, covering the State of Pernambuco, Brazil, and encompassing both tropical and semi-arid environments associated with the Atlantic Forest biomes and Caatinga;
2. The inclusion of water bodies annotated through expert-driven interpretation;
3. The representation of diverse hydrological contexts, including natural features such as lakes and lagoons, artificial structures such as reservoirs, dams, and small impoundments, as well as watercourses and coastal environments, characterized by strong spectral variability across space and time; and
4. The provision of a challenging benchmark dataset that captures a wide range of water body shapes, sizes, and environmental conditions, including complex scenarios such as urban areas and densely vegetated regions, thereby increasing the difficulty and relevance of the segmentation task.

2. Related Works

Several remote sensing image datasets used as benchmarks for evaluating neural networks are publicly available. Among the most commonly used datasets for water body segmentation are the Gaofen Image Dataset (GID) (Tong et al., 2020), the Land-Cover Domain Adaptive Semantic Segmentation (LoveDA) dataset (Wang et al., 2021), and Sen1Floods11 (Bonafilia et al., 2020).

The GID dataset consists of 150 Gaofen-2 images acquired over more than 60 cities in China, with dimensions of approximately 6800 × 7200 pixels and a spatial resolution of 4 m. The LoveDA dataset contains images from Google Earth, including urban and rural scenes from three Chinese cities, with a size of 1024 × 1024 pixels and a spatial resolution of 0.3 m. However, both datasets were originally designed for land-use and land-cover segmentation, often requiring sample selection and adaptation to a binary classification problem (water vs. non-water), which limits the number of usable samples.

Sen1Floods11 comprises Sentinel-1 images for mapping permanent water bodies and flood events, with patches of 512 × 512 pixels and a spatial resolution of 10 m. Similarly, other flood segmentation datasets have been proposed, such as FloodNet (Rahnemoonfar et al., 2021), based on high-resolution UAV imagery (1.5 cm); GF-FloodNet (Zhang et al., 2023), composed of Gaofen-3 and Gaofen-2 imagery; and UrbanSARFloods, (Zhao et al., 2024), which uses Sentinel-1 data and interferometric coherence, containing 8,879 samples of 512 × 512 pixels across five continents and 18 flood events. In contrast, most of these datasets are primarily focused on flood-related scenarios, limiting their applicability to general water body segmentation tasks.

Other datasets include LandCover.ai (Boguszewski et al., 2021) and DeepGlobe (Demir et al., 2018), designed for land cover segmentation including water classes; PakSAT (Abid et al., 2021) , developed using Sentinel-2 imagery for water segmentation; the Irish Coastal Segmentation (LICS) dataset (O’Sullivan et al., 2024) and the Sea-Land dataset (Yang et al., 2020) , both based on Landsat imagery for coastal water segmentation; and S1S2-Water (Wieland et al., 2024), a global dataset for semantic segmentation of surface water bodies using Sentinel-1 and Sentinel-2 data. Finally, SEN12-WATER (Russo et al., 2025) is a multi-source and multi-temporal dataset that integrates SAR data, elevation, slope, and multispectral optical bands acquired over time.

3. Method

3.1 Dataset

The Pernambuco Waterbody Dataset (PWD) is a multi-source remote sensing dataset that combines aerial and orbital multispectral imagery with pixel-level manual annotations for water body segmentation. This subsection describes the data used in this study and the methodology adopted for dataset construction, including data sources, coverage area, the construction process, and dataset organization.

3.1.1 Coverage Area: The images that compose the dataset cover the State of Pernambuco, Brazil (Figure 1). The region includes water bodies of varying shapes and sizes, such as natural lakes, reservoirs, and dams, and is composed of 13 river basins (APAC, 2023), with an emphasis on the São Francisco River and its tributaries, located in the southern part of the state.

The state is subdivided into three distinct mesoregions. The Zona da Mata, located in the eastern portion along the coastline, is characterized by high rainfall levels, with annual accumulated precipitation reaching approximately 2,000 mm (APAC, 2023), the presence of perennial rivers, and the predominance of the Atlantic Forest biome. The Agreste represents a transition zone between the Caatinga and Atlantic Forest biomes, with more humid conditions in its eastern portion and progressively drier conditions toward the interior, with periodic droughts. The Sertão, located in the western region, has a semi-arid climate, with irregular and scarce rainfall, annual accumulated precipitation ranging from 250 to 600 mm (APAC, 2023), predominantly intermittent rivers, and vegetation typical of the Caatinga biome.

Due to the region's environmental characteristics, especially in semi-arid areas, water bodies play a fundamental role in water supply, energy production, and agricultural irrigation.

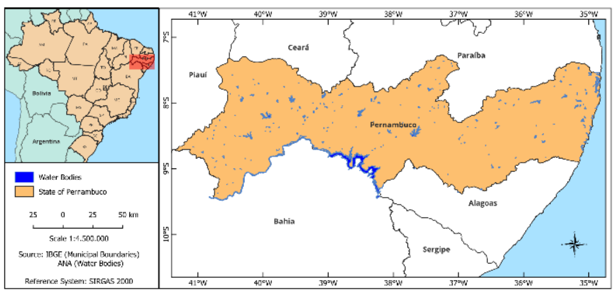

Figure 1. Coverage Area (State of Pernambuco).

3.1.2 Data Sources: Orbital images used in this study were obtained from the MultiSpectral Instrument (MSI) sensor onboard the Sentinel-2A satellite. According to the Copernicus Program, Sentinel-2A was launched in 2015 by the European Space Agency and is part of the Sentinel-2 mission, which also includes the Sentinel-2B satellite, launched in 2017. The MSI sensors have 13 spectral bands with different spatial resolutions (10, 20, and 60 m), a swath width of 290 km, and, together, provide a revisit time of approximately 5 days, enabling the monitoring and analysis of water bodies at large spatial and temporal scales.

The images were acquired through the Copernicus Data Space Ecosystem Browser, covering the period from January 2016 to January 2026, with maximum cloud coverage of 0.5% and processing level 2A, which provides orthorectified bottom-of-atmosphere (BOA) surface reflectance (Sentinel Hub, 2025).

Aerial images were obtained from the Pernambuco Tridimensional Program (PE3D), an initiative of the Government of the State of Pernambuco and the Secretariat of Economic Development. The program provides statewide aerial photogrammetric and LiDAR data covering the entire State of Pernambuco, Brazil, acquired between April 2014 and March 2017, totaling approximately 10,500 RGB orthophotos with a spatial resolution of 50 cm. The data are freely available at *https://www.pe3d.pe.gov.br/mapa.php.*

3.1.3 Construction: The construction process of the PWD was structured into sequential stages, including image preprocessing, water body vectorization, segmentation mask generation, image tiling into patches, and data splitting into training, validation, and test sets.

For the first version of the dataset, 517 aerial images with approximate dimensions of 7,119 × 4,856 pixels and 40 Sentinel-2A images with dimensions of approximately 10,980 × 10,980 pixels were selected. The scenes were manually selected to maximize the inclusion of water bodies, covering targets with diverse appearances, sizes, and shapes, as well as variations across environmental and geographic contexts, including urban, rural, tropical, and semi-arid regions, to enhance dataset representativeness and suitability for segmentation tasks.

Preprocessing of Sentinel-2A imagery was performed using a Python-based workflow, primarily relying on the Rasterio and NumPy libraries. Rasterio was used for reading, writing, and handling raster data, while NumPy was employed for matrix operations and spectral index calculations.

The selected bands for dataset composition were Blue (458–523 nm), Green (543–578 nm), Red (650–680 nm), and Near-infrared (NIR) (785–900 nm), all with a spatial resolution of 10 m. This selection is justified by the ability of visible bands (RGB) to capture spectral information related to color, texture, and surface context, while the NIR band is particularly effective for discriminating between water and land, as water strongly absorbs energy in this spectral range (Jensen, 2014).

The orbital images were individually reprojected to the SIRGAS 2000 coordinate reference system, using UTM zones 24S (EPSG:31984) and 25S (EPSG:31985), according to the spatial coverage of each scene. The nearest neighbor resampling method was applied to preserve the original reflectance values.

Subsequently, the Normalized Difference Water Index (NDWI) (McFeeters, 1996), was computed, as defined in Equation 1:

$$NDWI = \frac{Green - NIR}{Green + NIR} \tag{1}$$

The spectral bands and the NDWI were exported as separate GeoTIFF image channels. Aerial images did not undergo preprocessing, as they are already provided in RGB composition and referenced to the SIRGAS 2000 coordinate system, using UTM zones 24S (EPSG:31984) and 25S (EPSG:31985).

After preprocessing and scene selection, the manual annotation process was initiated, in which water bodies were vectorized by three cartographic experts through on-screen visual interpretation using the QGIS software (Long-Term Release version 3.40.7). Vectorization of aerial imagery was based on true-color (RGB) composition, whereas for Sentinel-2A imagery, water body identification and delineation were guided by the NDWI index. The annotation process was conducted independently to minimize potential bias in delineating water bodies. After the initial annotation, the results were reviewed to ensure consistency and reliability across the dataset.

The vectorization process considered spectral, geometric, and contextual criteria and was performed at a scale compatible with the spatial resolution of the images. To support feature identification, official vector datasets of water bodies from the National Water and Sanitation Agency and the 3rd Geoinformation Center were used.

It is important to note that, due to the environmental characteristics of the study area, where water bodies exhibit significant seasonal variations between dry and wet periods, the delineations performed directly on the images were adopted as ground truth, representing the conditions at the time of image acquisition.

Reference masks were generated from vector polygons of water bodies and converted to raster format using the same grid as the input images, yielding pixel-level binary masks, where 1 represents water and 0 represents non-water.

After mask generation, both images and their corresponding labels were partitioned into non-overlapping patches of 512 × 512 pixels. During this process, patches containing NoData values were automatically discarded.

Finally, the dataset was split into training, validation, and test sets in proportions of 60%, 20%, and 20%, respectively, ensuring no spatial overlap between subsets. The split was performed at the image level, so that Sentinel-2A images acquired at different dates from the same area were assigned to the same subset. This strategy aimed to prevent spatial leakage and to enable evaluation of the model’s generalization capability on unseen geographic areas.

Additionally, the split was designed to maintain a balanced distribution of water and non-water pixels across the subsets, ensuring consistent class representation across the training, validation, and test sets. Figure 2 presents the resulting distribution.

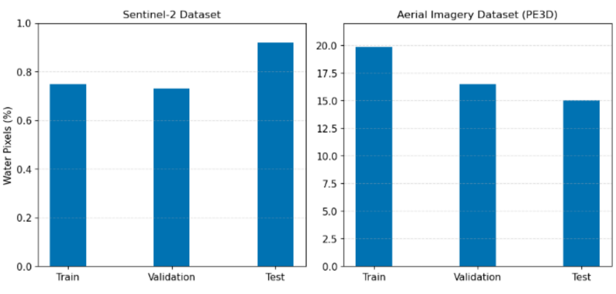

Figure 2. Water Pixel Distribution Across Dataset Splits.

A significant difference in the proportion of the water class is observed across the dataset subsets. While the subset derived from Sentinel-2A imagery contains approximately 0.78% water pixels, the aerial imagery subset presents around 19%. Sentinel-2A images cover large areas, predominantly composed of terrestrial surfaces, which reduces the relative proportion of water pixels and makes it more challenging to select representative regions of water bodies. In contrast, aerial imagery, due to its smaller spatial coverage and higher spatial detail, enables a more precise selection of regions with a higher water body presence.

The dataset includes structured metadata to support reproducibility and facilitate its use in further research. The metadata include information on geographic location, the proportion of water pixels per sample, data type and spectral bands, patch dimensions, and dataset partition (training, validation, or test).

In summary, the first version of the PWD consists of multi-source remote sensing image samples, with spatial resolutions of 0.5 m for aerial imagery and 10 m for Sentinel-2A orbital imagery. Each sample corresponds to 512 × 512 pixel patches, totalling 51,743 aerial image patches and 15,321 orbital image patches (Table 1). All samples and their corresponding labels include geographic coordinates and are stored as GeoTIFFs. It is important to note that the original spatial resolution of the images was preserved without resampling, to avoid degradation or loss of relevant spatial information

Source	PE3D	Sentinel – 2A
Platform	Airborne	Satellite
Spatial Resolution (m)	0.5	10
Bands	RGB	RGB + NIR
Number of Patches (512×512)	51,743	15,321

Table 1. Summary of the PE3D and Sentinel-2A Datasets

3.2 Models

To validate the applicability of the dataset for water body extraction, the U-Net model (Ronneberger et al., 2015), with ResNet34 and EfficientNet-B0 backbones; DeepLabV3+ (Chen et al., 2018), with ResNet50; and U-Net++ (Zhou et al., 2018), with ResNet34 were employed. These architectures are widely used in the literature for semantic segmentation tasks. All models adopt an encoder–decoder structure.

U-Net employs skip connections between corresponding encoder and decoder layers, allowing the network to retain high-frequency information during upsampling, thereby improving reconstruction accuracy and mitigating information loss and vanishing gradients (Sim et al., 2024; Wang et al., 2024). In U-Net++, in addition to skip connections, intermediate dense connections are introduced to facilitate semantic fusion between feature maps at different levels, reducing the semantic gap between the encoder and decoder (Lyu et al., 2023). The DeepLabV3+ model expands the receptive field through dilated convolutions and combines multi-scale feature maps using the Atrous Spatial Pyramid Pooling (ASPP) module (Chen et al., 2025).

The backbones were selected among those most commonly reported in the literature for water segmentation tasks, also considering computational resource constraints. ResNet (He et al., 2016), by stacking residual blocks, enables more effective capture of both local details and contextual information, enhancing the representational power of the network (Liu et al., 2024). EfficientNet, on the other hand, employs Mobile Inverted Bottleneck Convolution (MBConv) blocks combined with Squeeze-and-Excitation (SE) modules, and adopts a compound scaling strategy, enabling hierarchical feature extraction and the simultaneous capture of fine details and high-level contextual information, resulting in improved generalization capability (Gautam and Singhai, 2024).

3.3 Training Setup

The experiments were conducted in the Google Colab environment, using the PyTorch framework with NVIDIA A100 GPU acceleration. The segmentation architectures were implemented using the segmentation_models.pytorch library (Iakubovskii, 2019). The backbones were initialized with weights pre-trained on the ImageNet dataset.

Training was performed with a batch size of 64 images and a spatial resolution of 512 × 512 pixels. For aerial data, visible spectrum bands (RGB) were used, whereas for Sentinel-2A data, the near-infrared (NIR) band was also included.

The networks were trained for up to 100 epochs using the AdamW optimizer, with an initial learning rate of 1×10^{-4}. To address class imbalance, a weighted combination of Binary Cross-Entropy (BCE) and Dice Loss functions was employed.

To ensure reproducibility, a fixed random seed was adopted. A ReduceLROnPlateau scheduler was used to adjust the learning rate based on validation IoU, with a factor of 0.5, patience of 4 epochs, and a threshold of 0.002. Early stopping was applied with a patience of 10 epochs and a minimum IoU improvement of 0.002 after a 5-epoch warm-up period.

During training, a data augmentation strategy was applied to increase the size and diversity of the training set and reduce overfitting, thereby improving model generalization (Buslaev et al., 2020). Geometric and radiometric transformations were implemented using the Albumentations library (Buslaev et al., 2020). Geometric transformations included horizontal and vertical flips (HorizontalFlip and VerticalFlip) and rotations in multiples of 90° (RandomRotate90), applied probabilistically. Radiometric transformations involved random brightness and contrast adjustments (RandomBrightnessContrast) and the addition of Gaussian noise (GaussNoise), applied only to aerial images to simulate illumination variations and noise commonly found in remote sensing data.

3.4 Model evaluation metrics

In this study, four evaluation metrics were used: Recall, Precision, F1-Score and Intersection over Union (IoU). Recall (R) (Equation 2) is defined as the ratio between the number of correctly predicted pixels (True Positives) and the total number of positive pixels (Wang et al., 2020). Precision (P) (Equation 3) is the proportion of true positives among all predicted positive pixels. In other words, precision represents the proportion of pixels correctly classified as water among all pixels classified as water (Isikdogan et al., 2017). The F1-score (Equation 4) corresponds to the harmonic mean of Precision and Recall, ranging from 0 to 1, where 1 indicates perfect performance and 0 the worst possible performance (Li et al., 2019). The Intersection over Union (IoU) (Equation 5) also ranges from 0 to 1, where 0 indicates no overlap and 1 represents perfect agreement between the segmentation and the ground truth. Therefore, models with values closer to 1 perform better at identifying and classifying relevant pixels, ensuring that the segmented output closely matches the ground truth (Thomas Ramos and Sappa, 2024).

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

$$Precision\ (P) = \frac{TP}{TP + FP} \tag{3}$$

$$F1 - score\ = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

$$IoU = \frac{[A \cap B]}{|A \cup B|} = \frac{TP}{TP + FP + FN} \tag{5}$$

For the task of water body segmentation, True Positive (TP) corresponds to water pixels present in the input image that were correctly identified by the model as belonging to the water class; True Negative (TN) refers to non-water pixels correctly classified as non-water; False Positive (FP) occurs when non-water pixels are incorrectly classified as water (commission error); and False Negative (FN) corresponds to water pixels that were incorrectly classified as non-water (omission error). The best model evaluated in this study was selected based on the IoU metric on the validation set and subsequently used for evaluation on the test set.

4. Results

The models were trained separately for each data source, using visible spectrum bands (RGB) for aerial imagery and both RGB and near-infrared (NIR) bands for Sentinel-2A imagery.
Table 2 shows that, among the evaluated architectures, U-Net++ with a ResNet34 encoder achieved the best performance in the segmentation of aerial imagery, reaching an IoU of 0.946 and an F1-score of 0.972 on the test set, demonstrating the model’s ability to accurately identify water pixels. The remaining architectures also achieved similar results, with values above 0.92 and differences below 0.02 across all metrics, indicating consistent performance among the evaluated models.

Architecture	Backbone	P	R	F1-score	IoU
U-Net	ResNet34	0.967	0.976	0.971	0.944
U-Net	EfficientNet-b0	0.959	**0.979**	0.969	0.939
DeepLabV3+	ResNet50	0.967	0.974	0.970	0.943
U-Net++	ResNet34	**0.971**	0.974	**0.972**	**0.946**

Table 2. Evaluation metrics for water body segmentation on aerial imagery

It is important to note that U-Net++ presented a lower recall (0.974) but achieved the highest precision (0.971), whereas the EfficientNet-b0 architecture obtained the highest recall (0.979) and the lowest precision (0.959). In this context, U-Net++ achieves the best F1-score (0.972), indicating a more balanced trade-off between precision and recall, which reflects a better balance between omission and commission errors, especially in spectrally complex regions.
The model inferences across different geographic contexts, including urban and rural environments, spanning semi-arid and tropical regions, and covering water bodies with varying sizes, shapes, and tonalities, are presented in Figure 3. The scenes include: (a) Itamaragi reservoir, Exu–PE; (b) a reservoir located in a semi-arid region, Granito–PE; (c) a reservoir in the Zona da Mata region, Barra de Sirinhaém–PE; (d) margins of the São Francisco River; (e) a tributary of the Capibaribe River, Recife–PE; (f) Botafogo Dam, Igarassu–PE; and (g) a reservoir in Itapissuma–PE.

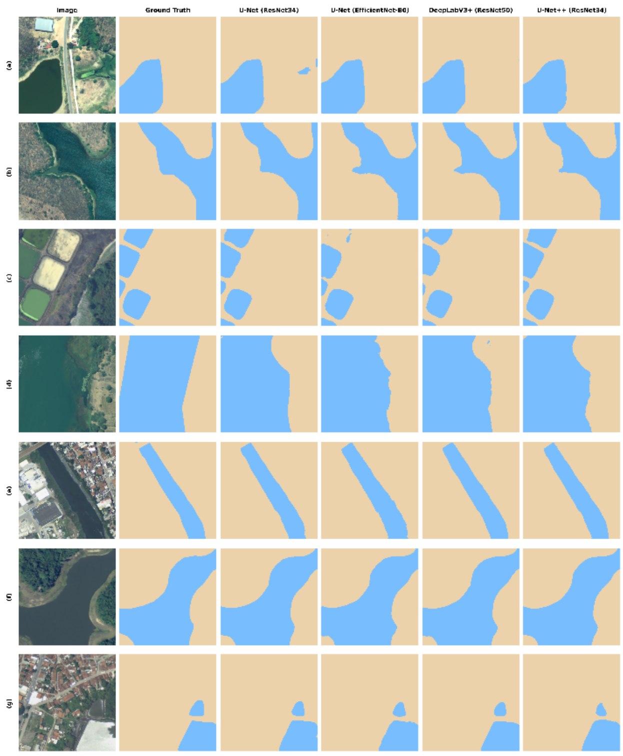

Figure 3. Water segmentation results using aerial imagery

The results show segmentation masks with well-defined water boundaries, with only minor inaccuracies at object edges, low noise levels, and a reduced occurrence of false positives.
The results obtained from the models for Sentinel-2A image segmentation (Table 3) indicate that the U-Net with a ResNet34 backbone achieved the best performance, reaching an IoU of 0.871 and an F1-score of 0.931 on the test set. Similar to the results observed for aerial imagery, the models exhibited comparable performance, with only minor variations across the evaluated metrics. Notably, U-Net++ with ResNet34 achieved a precision of 0.956, while DeepLabV3+ with ResNet50 achieved a recall of 0.930.

Architecture	Backbone	P	R	F1-score	IoU
U-Net	ResNet34	0.943	0.919	**0.931**	**0.871**
U-Net	EfficientNet-b0	0.942	0.914	0.928	0.865
DeepLabV3+	ResNet50	0.923	**0.930**	0.926	0.863
U-Net++	ResNet34	**0.956**	0.893	0.924	0.858

Table 3. Evaluation metrics for water body segmentation on Sentinel-2A imagery.

As observed in the aerial data, U-Net++ achieves the highest precision (0.956) but the lowest recall (0.893), whereas DeepLabV3+ presents the highest recall (0.930) but the lowest precision (0.923). However, U-Net provides the best balance between both metrics, achieving the highest F1-score (0.931).
Similarly, the segmentation results for Sentinel-2A imagery are presented in Figure 4. The scenes include: (a) margins of the São Francisco River; (b) Ingazeira Dam, Ingazeira–PE; (c) Poções reservoir, Monteiro–PB; (d) Algodões reservoir, Ouricuri–PE; (e) São Francisco River, Petrolina–PE; (f) Poço da Cruz reservoir, Ibimirim–PE; and (g) Luanda stream, Serra Talhada–PE.

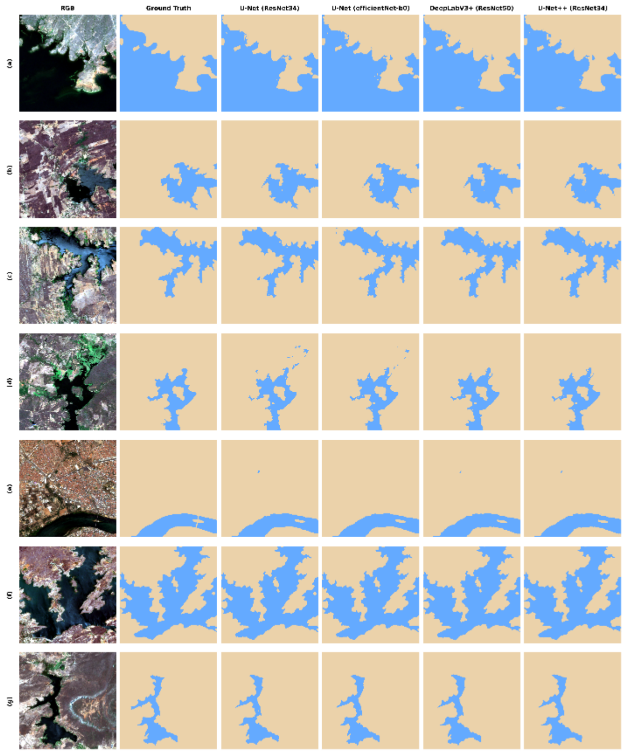

Figure 4. Water segmentation using Sentinel-2A imagery.

The segmentation masks show coherent delineation of water features, with low noise levels and a reduced occurrence of false positives. However, less smooth boundaries and some inaccuracies are observed, as well as partial omissions of narrower features in certain cases, compared to the aerial imagery results.

Overall, the models demonstrate satisfactory generalization, identifying water bodies of varying sizes, shapes, and spectral patterns. However, both experiments showed greater difficulty in segmenting areas with partial vegetation cover and narrow rivers.

The performance difference observed between aerial and Sentinel-2A imagery is mainly attributable to spatial resolution, particularly in accurately delineating water body boundaries. The higher spatial resolution of aerial data (0.5 m) enables better representation of fine structures, such as narrow rivers and complex boundaries, whereas in Sentinel-2A imagery (10 m), water body boundaries are often not clearly defined and are inaccurately labeled, frequently being represented by mixed pixels, which leads to discrepancies in the predictions, especially in boundary regions.

Additionally, the subset derived from Sentinel-2A imagery contains approximately 0.78% water pixels, whereas the aerial imagery subset presents around 19%. The reduced class imbalance in the aerial dataset likely facilitated model convergence and resulted in improved segmentation performance.

5. Conclusion

In this study, the Pernambuco Waterbody Dataset (PWD) was developed, a multi-source dataset composed of aerial and orbital remote sensing imagery designed for water body segmentation tasks. The dataset construction process included data acquisition, preprocessing, manual annotation, vectorization, segmentation mask generation, image tiling into patches, and data partitioning into training, validation, and test sets.

The dataset was validated using deep neural networks based on encoder–decoder architectures, with different backbone networks evaluated in the experiments. Among the evaluated configurations, ResNet34 consistently appears as the backbone of the top-performing models in both the PE3D and Sentinel-2A experiments, whereas more complex configurations, such as EfficientNet-b0 and DeepLabV3+ with ResNet50, did not achieve the best overall results.

This suggests that, within the context of this benchmark, increased architectural complexity does not necessarily result in proportional performance gains. In other words, lightweight and well-established architectures can provide competitive performance while maintaining lower computational cost and greater efficiency.

It was observed that the segmentation performance of Sentinel-2A imagery (RGB+NIR) was significantly lower than that obtained with high-resolution aerial RGB data, even with the inclusion of the additional NIR band. This result indicates that, within the context of this benchmark, spatial resolution and class distribution have a greater influence on model performance than the incorporation of additional spectral information.

The evaluation metrics and results indicate that the dataset is consistent and suitable for training semantic segmentation models. The first version of the Pernambuco Waterbody Dataset (PWD) will be publicly released upon publication of this work, incorporating the improvements identified in this study. The dataset will be made available in accordance with the licensing terms of the original data sources (Sentinel-2A and PE3D), including the imagery, corresponding segmentation masks, and associated metadata.

Future work includes revising and refining parts of the annotations to improve target delineation and incorporate previously unlabeled water bodies, aiming to enhance segmentation accuracy. Additionally, the incorporation of Digital Elevation Model (DEM) data for aerial imagery and the NDWI index for Sentinel-2A imagery is planned, which may improve water body discrimination. Finally, future investigations will explore more recent architectures, such as Transformers or hybrid approaches (CNN + Transformer).

References

Abid, N., Shahzad, M., Malik, M.I., Schwanecke, U., Ulges, A., Kovács, G., Shafait, F., 2021. UCL: Unsupervised Curriculum Learning for water body classification from remote sensing imagery. International Journal of Applied Earth Observation and Geoinformation 105, 102568. https://doi.org/10.1016/J.JAG.2021.102568

Agência Nacional de Águas e Saneamento Básico (ANA), 2022. ANA aumenta em cerca de quatro vezes sua base de dados sobre massas d’água do Brasil.

Agência Pernambucana de Águas e Clima (APAC), 2023. Atlas climatológico do Estado de Pernambuco: normais climatológicas 1991–2020. Recife, Brasil.

Attya, M., Abo-Seida, O.M., Abdulkader, H.M., Mohammed, A.M., 2025. A hybrid deep learning approach for accurate water body segmentation in satellite imagery. Earth Sci. Inform. 18, 1–29. https://doi.org/10.1007/S12145-025-01912-Y/FIGURES/17

Bangira, T., Alfieri, S.M., Menenti, M., van Niekerk, A., 2019. Comparing thresholding with machine learning classifiers for mapping complex water. Remote Sens. (Basel). 11. https://doi.org/10.3390/RS11111351

Boguszewski, A., Batorski, D., Ziemba-Jankowska, N., Dziedzic, T., Zambrzycka, A., 2021. LandCover.ai: Dataset for automatic mapping of buildings, woodlands, water and roads from aerial imagery. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 1102–1110. https://doi.org/10.1109/CVPRW53098.2021.00121

Bonafilia, D., Tellman, B., Anderson, T., Issenberg, E., 2020. Sen1Floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 2020-June, 835–845. https://doi.org/10.1109/CVPRW50498.2020.00113

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A., 2020. Albumentations: Fast and Flexible Image Augmentations. Information 2020, Vol. 11, 11. https://doi.org/10.3390/info11020125

Chen, B., Zhang, Q., Yang, N., Wang, X., Zhang, X., Chen, Y., Wang, S., 2025. Deep Learning Extraction of Tidal Creeks in the Yellow River Delta Using GF-2 Imagery. Remote Sensing 2025, Vol. 17, Page 676 17, 676. https://doi.org/10.3390/RS17040676

Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11211 LNCS, 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

Chen, X., Pan, H., Liu, J., 2024. SwinDefNet: A Novel Surface Water Mapping Model in Mountain and Cloudy Regions Based on Sentinel-2 Imagery. Electronics 2024, Vol. 13, Page 2870 13, 2870. https://doi.org/10.3390/ELECTRONICS13142870

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., Hughes, F., Tuia, D., Raskar, R., 2018. Deepglobe 2018: A challenge to parse the earth through satellite images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 172–181.

Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X., 2016. Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band. Remote Sens. (Basel). 8, 354. https://doi.org/10.3390/RS8040354

Gautam, S., Singhai, J., 2024. Critical review on deep learning methodologies employed for water-body segmentation through remote sensing images. Multimed. Tools Appl. 83, 1869–1889. https://doi.org/10.1007/S11042-023-15764-5

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-December, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hibjur Rahaman, M., Roshani, Masroor, M., Sajjad, H., 2023. Integrating remote sensing derived indices and machine learning algorithms for precise extraction of small surface water bodies in the lower Thoubal river watershed, India. J. Clean. Prod. 422, 138563. https://doi.org/10.1016/J.JCLEPRO.2023.138563

Iakubovskii, P., 2019. Segmentation Models Pytorch. GitHub repository.

Isikdogan, F., Bovik, A.C., Passalacqua, P., 2017. Surface water mapping by deep learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 4909–4918. https://doi.org/10.1109/JSTARS.2017.2735443

Jensen, J.R.., 2014. Remote sensing of the environment : an earth resource perspective. Pearson.

Jonnala, N.S., Bheemana, R.C., Prakash, K., Bansal, S., Jain, A., Pandey, V., Faruque, M.R.I., Al-mugren, K.S., 2025. DSIA U-Net: deep shallow interaction with attention mechanism UNet for remote sensing satellite images. Sci. Rep. 15. https://doi.org/10.1038/S41598-024-84134-4

Klein, I., Gessner, U., Dietz, A.J., Kuenzer, C., 2017. Global WaterPack – A 250 m resolution dataset revealing the daily dynamics of global inland water bodies. Remote Sens. Environ. 198, 345–362. https://doi.org/10.1016/J.RSE.2017.06.045

Knight, S.J., McClean, C.J., White, P.C.L., 2022. The importance of ecological quality of public green and blue spaces for subjective well-being. Landsc. Urban Plan. 226, 104510. https://doi.org/10.1016/J.LANDURBPLAN.2022.104510

Lei, P., Yi, J., Li, S., Li, Y., Lin, H., 2025. Agricultural surface water extraction in environmental remote sensing: A novel semantic segmentation model emphasizing contextual information enhancement and foreground detail attention. Neurocomputing 617, 129110. https://doi.org/10.1016/J.NEUCOM.2024.129110

Li, A., Fan, M., Qin, G., Xu, Y., Wang, H., 2021. Comparative analysis of machine learning algorithms in automatic identification and extraction of water boundaries. Applied Sciences (Switzerland) 11, 10062. https://doi.org/10.3390/APP112110062/S1

Li, B., Xie, Y., Tao, Z., Shao, W., Qin, G., 2025. Refined monitoring of water bodies in built-up areas and analysis and evaluation of water color characteristics in complex urban backgrounds. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. https://doi.org/10.1109/JSTARS.2025.3575847

Li, J., Meng, Y., Li, Y., Cui, Q., Yang, X., Tao, C., Wang, Z., Li, L., Zhang, W., 2022. Accurate water extraction using remote sensing imagery based on normalized difference water index and unsupervised deep learning. J. Hydrol. (Amst). 612, 128202. https://doi.org/10.1016/J.JHYDROL.2022.128202

Li, L., Yan, Z., Shen, Q., Cheng, G., Gao, L., Zhang, B., 2019. Water Body Extraction from Very High Spatial Resolution Remote Sensing Data Based on Fully Convolutional Networks. Remote Sensing 2019, Vol. 11, Page 1162 11, 1162. https://doi.org/10.3390/RS11101162

Liu, M ;, Liu, J ;, Hu, H.A., Liu, Min, Liu, Jiangping, Hu, H., 2024. A Novel Deep Learning Network Model for Extracting Lake Water Bodies from Remote Sensing Images. Applied Sciences 2024, Vol. 14, Page 1344 14, 1344. https://doi.org/10.3390/APP14041344

Lyu, X., Jiang, W., Li, X., Fang, Y., Xu, Z., Wang, X., 2023. MSAFNet: Multiscale Successive Attention Fusion Network for Water Body Extraction of Remote Sensing Images. Remote Sensing 2023, Vol. 15, Page 3121 15, 3121. https://doi.org/10.3390/RS15123121

McFeeters, S.K., 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. Int. J. Remote Sens. 17, 1425–1432. https://doi.org/10.1080/01431169608948714

Nagaraj, R., Kumar, L.S., 2023. Surface water body extraction and Change Detection Analysis using Machine Learning Algorithms: A Case study of Vaigai Dam, India. Proceedings of 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication, IConSCEPT 2023. https://doi.org/10.1109/ICONSCEPT57958.2023.10170342

O’Sullivan, C., Kashyap, A., Coveney, S., Monteys, X., Dev, S., 2024. Enhancing coastal water body segmentation with Landsat Irish Coastal Segmentation (LICS) dataset. Remote Sens. Appl. 36, 101276. https://doi.org/10.1016/J.RSASE.2024.101276

Portalés-Julià, E., Mateo-García, G., Purcell, C., Gómez-Chova, L., 2023. Global flood extent segmentation in optical satellite images. Scientific Reports 2023 13:1 13, 1–14. https://doi.org/10.1038/s41598-023-47595-7

Rahnemoonfar, M., Chowdhury, T., Sarkar, A., Varshney, D., Yari, M., Murphy, R.R., 2021. FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding. IEEE Access 9, 89644–89654. https://doi.org/10.1109/ACCESS.2021.3090981

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9351, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Russo, L., Mauro, F., Sebastianelli, A., Gamba, P., Ullo, S.L., 2025. An End-to-End Framework for Multimodal and Multitemporal Analysis of Water Bodies Using the SEN12-WATER Dataset. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 18, 27615–27628. https://doi.org/10.1109/JSTARS.2025.3622914

Sentinel Hub, 2025. Copernicus Data Space Ecosystem Browser.

Sim, W.D., Yim, J.S., Lee, J.S., 2024. Assessing Land Cover Classification Accuracy: Variations in Dataset Combinations and Deep Learning Models. Remote Sensing 2024, Vol. 16, Page 2623 16, 2623. https://doi.org/10.3390/RS16142623

Thomas Ramos, L., Sappa, A.D., 2024. Multispectral Semantic Segmentation for Land Cover Classification: An Overview. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 14295–14336. https://doi.org/10.1109/JSTARS.2024.3438620

Tong, X.Y., Xia, G.S., Lu, Q., Shen, H., Li, S., You, S., Zhang, L., 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sens. Environ. 237, 111322. https://doi.org/10.1016/J.RSE.2019.111322

Vongkusolkit, J., Peng, B., Wu, M., Huang, Q., Andresen, C.G., 2023. Near Real-Time Flood Mapping with Weakly Supervised Machine Learning. Remote Sensing 2023, Vol. 15, Page 3263 15, 3263. https://doi.org/10.3390/RS15133263

Wang, H., Zheng, Y., Chen, O., 2024. High-precision Qiantang River Water Body Recognition Based on Remote Sensing Image. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3510796

Wang, J., Zheng, Z., Ma, A., Lu, X., Zhong, Y., 2021. LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation, in: Vanschoren, J., Yeung, S. (Eds.), Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks.

Wang, Z., Gao, X., Zhang, Y., Zhao, G., 2020. MSLWENet: A novel deep learning network for lake water body extraction of Google remote sensing images. Remote Sens. 12, 4140. https://doi.org/10.3390/rs12244140

Wieland, M., Fichtner, F., Martinis, S., Groth, S., Krullikowski, C., Plank, S., Motagh, M., 2024. S1S2-Water: A Global Dataset for Semantic Segmentation of Water Bodies From Sentinel- 1 and Sentinel-2 Satellite Images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 1084–1099. https://doi.org/10.1109/JSTARS.2023.3333969

Wu, W., Li, Q., Zhang, Y., Du, X., Wang, H., 2018. Two-step urban water index (TSUWI): A new technique for high-resolution mapping of urban surface water. Remote Sens. 10, 1704. https://doi.org/10.3390/rs10111704

Xiang, Z., Jun, L.Y., Wei, R.T., Jiang, L.Y., 2025. MTAU-Net: multi-scale feature fusion network for water body segmentation. J. Appl. Remote Sens. 19, 21010. https://doi.org/10.1117/1.JRS.19.021010

Yang, T., Jiang, S., Hong, Z., Zhang, Y., Han, Y., Zhou, R., Wang, J., Yang, S., Tong, X., Kuc, T. yong, 2020. Sea-Land Segmentation Using Deep Learning Techniques for Landsat-8 OLI Imagery. Marine Geodesy 43, 105–133. https://doi.org/10.1080/01490419.2020.1713266

Yu, Y., Yao, Y., Guan, H., Li, D., Liu, Z., Wang, L., Yu, C., Xiao, S., Wang, W., Chang, L., 2021. A self-attention capsule feature pyramid network for water body extraction from remote sensing imagery. Int. J. Remote Sens. 42, 1801–1822. https://doi.org/10.1080/01431161.2020.1842544

Yuan, X., Shi, J., Gu, L., 2021. A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Syst. Appl. 169, 114417. https://doi.org/10.1016/J.ESWA.2020.114417

Zhai, K., Wu, X., Qin, Y., Du, P., 2015. Comparison of surface water extraction performances of different classic water indices using OLI and TM imageries in different situations. Geo-Spatial Information Science 18, 32–42. https://doi.org/10.1080/10095020.2015.1017911

Zhang, Q., Hu, X., Xiao, Y., 2023. A novel hybrid model based on CNN and multi-scale transformer for extracting water bodies from high resolution remote sensing images. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 10, 889–894. https://doi.org/10.5194/isprs-annals-x-1-w1-2023-889-2023

Zhang, Y., Liu, P., Chen, L., Xu, M., Guo, X., Zhao, L., 2023. A new multi-source remote sensing image sample dataset with high resolution for flood area extraction: GF-FloodNet. Int. J. Digit. Earth 16, 2522–2554. https://doi.org/10.1080/17538947.2023.2230978

Zhao, J., Xiong, Z., Zhu, X.X., 2024. UrbanSARFloods: Sentinel-1 SLC-Based Benchmark Dataset for Urban and Open-Area Flood Mapping. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 419–429. https://doi.org/10.1109/CVPRW63382.2024.00047

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 11045 LNCS, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1