International Journal of Aquatic Research and Environmental Studies 6(2) 727-754 2026

A Review of Surface Water Body Extraction from Remote Sensing Images: From Visual Interpretation to Transformer-Based Approaches

Santosh Onker[1]*, Alpana Pandey[2]

Department of Electronics and Communication Engineering, Maulana Azad National Institute of Technology, Bhopal 462003, Madhya Pradesh, India, Email: 2331401002@stu.manit.ac.in
Department of Electronics and Communication Engineering, Maulana Azad National Institute of Technology, Bhopal 462003, Madhya Pradesh, India, Email: alpanasubodh@gmail.corn

Abstract: The correct delineation of surface water bodies from remote sensing imagery (RSI) is essential for surface water monitoring, water conservation, and environmental change assessment. Various methods have been developed over the last three decades to improve the accuracy of surface water body extraction. This review provides a detailed overview of remote sensing-based surface water body extraction methods. Freely available datasets, pre-processing techniques, loss functions, and evaluation metrics, along with their advantages and limitations, are also discussed in this article. The review highlights a clear transition from traditional methods, such as visual interpretation and thresholding, toward machine learning (ML) methods and deep learning (DL) approaches, which include conventional deep learning models and hybrid transformer-based methods due to their superior ability to represent complex spatial and spectral properties of surface water bodies. Detection of small and fragmented water bodies, class imbalance, computational costs, model generalization, and the small size of annotated datasets are still significant challenges. This article summarizes recent advancements, current challenges, and future research directions. This article helps beginners and researchers understand key remote sensing concepts and their applications in surface water body extraction (SWBE).

Keywords: Satellite imagery, Surface water body extraction, Feature extraction, Convolutional Neural Network (CNN) , Machine learning & Deep learning models,Transformer-based models

Introduction

Surface water resources, such as rivers, lakes, reservoirs, wetlands, and coastal water bodies, are essential parts of the Earth's hydrological cycle and support ecological systems, farming, industry, hydropower generation, and domestic applications(Gleick, 1993). In recent decades, rapid urbanization, industrial development, climate change, and population growth have exerted considerable pressure on freshwater resources worldwide (Lee et al., 2023) (Vörösmarty et al., 2010). Moreover, the increasing occurrence of extreme hydrological events, such as floods, glacier retreat, sea-level rise, and drougbts, has highlighted the need for continuous monitoring and effective management of surface water resources (Hirabayashi et al., 2013). Accurate information on the spatial extent and temporal dynamics of surface water bodies is therefore vital for environmental monitoring, disaster mitigation, water resource planning, and sustainable development (Pekel et al., 2016). Traditional methods for monitoring surface water resources primarily depend on ground-based surveys and in-situ measurements. Thèse methods ensure reliable local-scale rneasurements but usually are laborious, tedious, and impractical for large-scale geographic regions. Consequently, remote sensing is widely used for large-scale surface water body monitoring because it provides regular, wide-area, and cost-effective observations (Lillesand, Kiefer and Chipinan, 2015). The availability of Eanh observation data acquired from satellites, airborne platforms, and unmanned aerial vehicles (UAVs) has significantly improved the ability to monitor hydrological systems on local, regional, and global scales (Wulder et al., 2019) (Drusch et al., 2012). Remote sensing refers to the acquisition of information about Earth's surface without direct physical contact with the target object (Campbell and Wynne, 2011). Satellites, aircraft, and UAVs are equipped with sensors that record electromagnetic (EM) radiation, either reflected or emitted, from different land cover classes, such as surface water bodies, vegetation, soil, forests, and urban infrastructure. The acquired information is subsequently processed and analyzed to generate imagery for environmental monitoring and resource management applications. Figure 1 shows the overall remote sensing process and how satellite images are generated. A key principle underlying remote sensing-based surface water body extraction (SWBE) is the interaction between EM radiation and surface materials. Since different land cover classes exhibit distinct spectral signatures, they can be identified and classified within remotely sensed imagery (Lillesand, Kiefer and Chipman, 2015) .

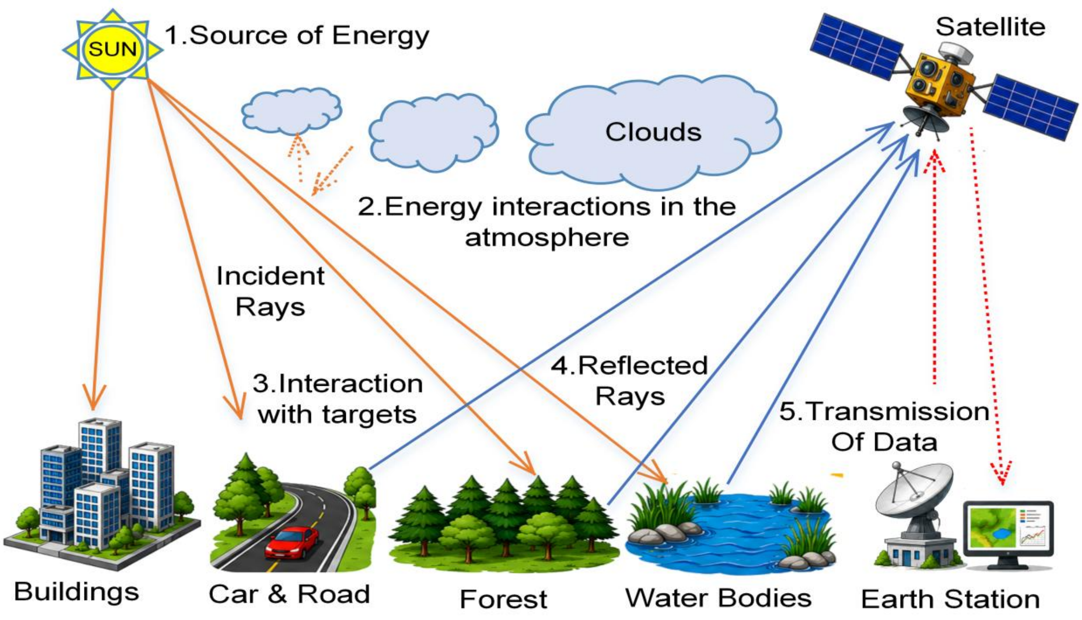

Figure 1: Schematic representation of the complete workflow of remote sensing image acquisition from satellite, including data capture, transmission, and processing

Water, vegetation, and bare soil exhibit distinct reflectance patterns across the electromagnetic (EM) spectrum (Bijeesh and Narasirnhamurthy, 2020). These patterns can be studied through reflectance curves. The reflectance curve is a graph plotted between the reflected energy and the wavelengths. Figure 2 illustrates the typical spectral reflectance curves for Earth's surface objects like water, soil, and vegetation. Water exhibits unique spectral properties in relation to other land surface types.

In the visible region of the EM spectrum, water generally exhibits low to moderate reflectance, depending on factors such as water depth, suspended sediments, chlorophyll concentration, and surface roughness (Mobley, 1994). However, in the near-infrared (NIR) and shortwave infrared (SWIR) regions, water strongly absorbs incident radiation and consequently exhibits very low reflectance values compared with vegetation, bare soil, and urban surfaces (McFEETERS, 1996) (Xu, 2006). Microwave signals also interact differently with water surfaces due to their dielectric propenies and surface roughness (Mason et al., 2009). Despite substantial advances in remote sensing technology, accurate surface water body extraction (SWBE) remains a challenging task (Ji, Zhang and Wylie, 2009). The spectral response of water varies due to seasonal fluctuations, sediment concentration, aquatic vegetation, water turbidity, and atmospheric effects (Feyisa et al., 2014). Furthermore, shadows, mountains, and buildings often exhibit spectral characteristics similar to water, leading to frequent mrs-classification errors (Li et al., 2013). Additional challenges arise from mixed pixels, narrow rivers, fragmented wetlands, complex urban environments, and the diversity of sensor characteristics and spatial resolutions (Pekel et al., 2016). The evolution of SWBE methods can be broadly categorized into visual interpretation methods, threshold-based approaches, spectral-index-based techniques, ML methods, and DL methods, including conventional DL models and hybrid transformer-based architectures (Gautam and Singhai, 2024). Figure 3 illustrates an overview of key methods employed for SWBE. Early surface water mapping studies primarily relied on manual visual interpretation of aerial photographs and satellite imagery, in which domain experts identified and delineated water bodies based on tone, texmre, shape, and contextual information (Lillesand, Kiefer and Chipman, 2015).

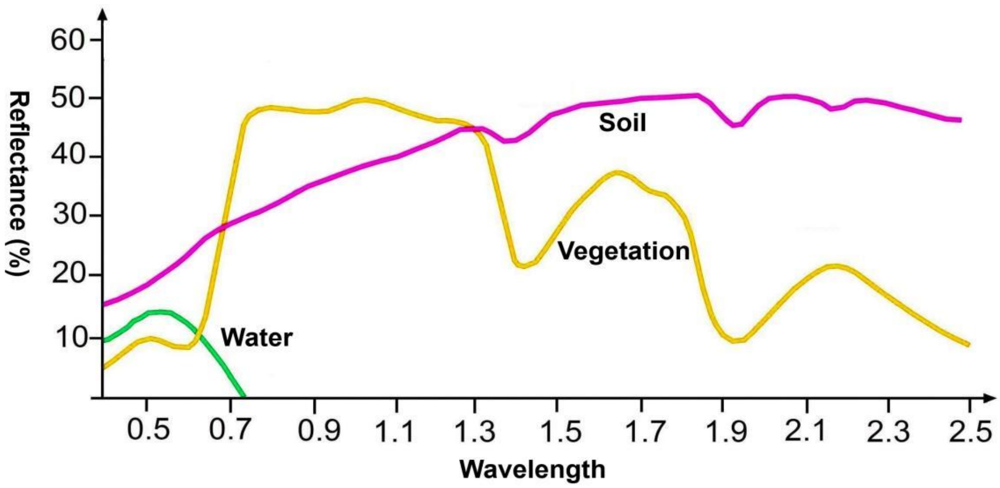

Figure 2: Spectral reflectance curves of water, soil, and vegetation across different wavelengths (Bijeesh and Narasimliamurthy, 2020) (Lillesand, Kiefer and Chipman, 2015)

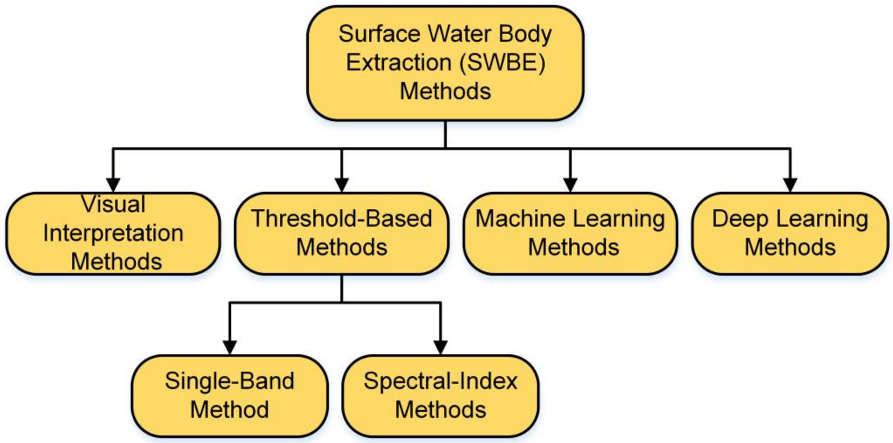

Figure 3: Taxonomy of surface water body extraction (SWBE) methods from remote sensing imagery

Although these approaches often produced satisfactory results, they were subjective, laborious, and unsuitable for large geographic areas.

Researchers then introduced threshold-based methods that classify water pixels using spectral reflectance values or image histogram characteristics (Ji, Zhang and Wylie, 2009). Techniques such as density slicing and Otsu thresholding became popular because of their simplicity and computational efficiency (Singh, 1989). However, these approaches have certain limitations in heterogeneous environments and are highly sensitive to threshold selection and image quality. The availability of multi-spectral satellite imagery further facilitated the development of spectral-index-based approaches. The most popular spectral index rnethod, the Normalized Difference Water Index (NDWI), was proposed by (McFEETERS, 1996). The author used the green and near-infrared (NIR) bands of Landsat TM imagery to identify surface water bodies. NDWI achieved a good classification accuracy for open surface water bodies by enhancing water features and suppressing vegetation and soil backgrounds. However, its performance degrades in complex environments, particularly in urban areas, where built-up surfaces often exhibit spectral responses similar to water, leading to misclassification.

To overcome the limitations of NDWI, (Xu, 2006) developed another spectral index method called the Modified Normalized Difference Water Index (MNDWI). The author replaced the NIR band with the shortwave infrared (SWIR) band and computed MNDWI using the green and SWIR bands of Landsat TM imagery. This modification effectively suppresses built-up land and enhances water features, resulting in improved water extraction accuracy in urban environments. However, MNDWI is still sensitive to complex environments, such as terrain shadows, cloud shadows, and highly heterogeneous land-cover conditions. To further address these limitations, (Feyisa et al., 2014) proposed the Automated Water Extraction Index (AWEI). The authors suggested two formulations: AWEI-sh for shadowed environments and AWEI-nsh for non-shadowed environments. Compared with NDWI and MNDWI, AWEI reduces the influence of shadows and dark surfaces, thereby improving water extraction accuracy in complex landscapes. However, its performance may still vary depending on sensor characteristics, regional conditions, and threshold selection.

Subsequently, (Wang et a1., 2015) proposed the Enhanced Water Index (EWI). The EWI was developed based on the MNDWI framework to improve the estimation of surface water fraction at the sub-pixel level using Landsat imagery. By enhancing the contrast between water and background features, EWI demonstrated improved capability for estimating mixed-water pixels and inland surface water features. Despite these improvements, spectral index-based methods remain sensitive to environmental variability, atmospheric effects, and mixed pixels.

To address these limitations, ML techniques such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees (DT), Artificial Neural Networks (ANN), and Randorn Forests (RF) came into the picture and were adopted for SWBE (Pal and Mather, 2005) (Belgiu and Drăguț, 2016). These ML models learn complex class boundaries beyond threshold-based methods. Supervised classifiers such as SVM, RF, and other ensemble methods have been used together with spectral indices such as NDWI, MNDWI, and AWEI, as well as textural features, to distinguish water from non-water. These approaches improve accuracy over single-index thresholding in heterogeneous landscapes but remain sensitive to feature design, hyper-parameter tuning, and mixed pixels (Belgiu and Drăguț, 2016). Recently, deep learning architectures such as Fully Convolutional Networks (Long, Shelhainer and Darrell, 2015), U-Net (Ronneberger, Fischer and Brox, 2015), SegNet (Badrinarayanan, Kendall and Cipolla, 2017), Deeplab (Chen et a1., 2017), and attention-based networks have enabled end-to-end semantic segmentation of water from multispectral and SAR imagery, providing state-of-the-art performance in complex environments with shadows, turbidiy, and heterogeneous backgrounds by automatically learning multi-scale spatial-spectral features.

Recent advances in computer vision have further introduced transforrner-based architectures into remote sensing applications. Models such as Vision Transformer (ViT) (Dosovitskiy et al., 2020), Swin Transformer (Lin et al., 2021), and SegFormer (Xie et al., 2021) achieve excellent performance by effectively modeling long-range contextual dependencies and global image relationships. In addition, hybrid CNN-transforiner architectures have ernerged as powerful alternatives that combine local feature extraction capabilities with global contextual

modeling (Dai et al., 2021). Despite these advances, multimodal data fusion and real-time monitoring remain key challenges in SWBE (Wulder et at., 2019).
Figure 4 presents the chronological evolution of major SWBE paradigms, highlighting the transition from conventional approaches to modern artificial intelligence-driven methods.

1.1 Satellite Platforms and Sensors

Recent developments in Earth observation technologies have greatly improved the ability to monitor surface water resources over diverse spatial and temporal scales. Various satellite platforms, such as Landsat (Land Satellite), Envisat (Environmental Satellite), and RADARSAT (Radar Satellite), equipped with different types of sensors, have been launched to acquire high-quality remote sensing data for environmental monitoring applications. These platforms provide observations with varying spatial, spectral, radiometric, and temporal resolutions (Wulder et al., 2019). Sensors used in remote sensing are categorized into passive and active classes based on their mode of operation. Passive sensors record naturally reflected or emitted EM radiation from Earth's surface, primarily originating from solar illumination (Lillesand, Kiefer and Chipman, 2015), whereas active sensors generate their own signal and send it towards the surface and record the back-reflected signal. These two sensors are usually called optical sensors and microwave or radar-based sensors.

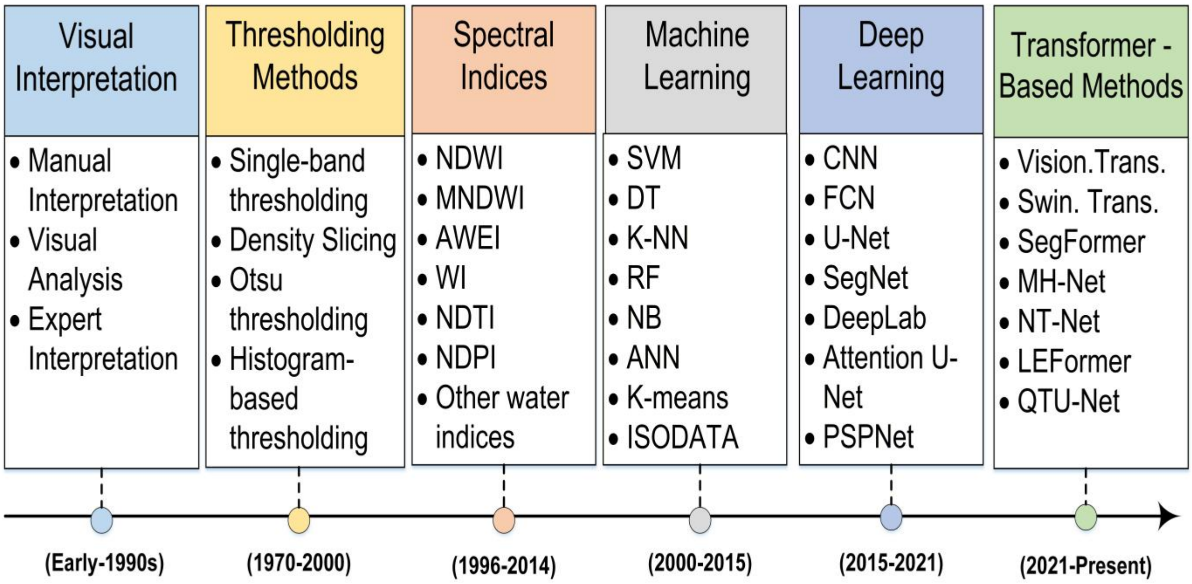

Figure 4:Chronological development of SWBE methods from remote sensing imagery

1.1.1 Optical Sensors

Optical sensors such as Multispectral Scanner System (MSS), Thematic Mapper(TM), Linear Imaging Self Scanning (LISS), Satellite Pour l'Observation de la Terre (SPOT), Enhanced Thematic Mapper (ETM+), Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), Advanced Very High Resolution Radioineter (AVHRR), Panchromatic Multi-spectral, and Moderate Resolution Imaging Spectroradioineter (MODIS) are available on various satellite platforms (Nath and Deb, 2010) lNagaraj and Kumar, 2024). This sensor acquires imagery in multiple spectral bands across the EM spectrum, including visible, near-infrared (NIR), and short-wave infrared (SWIR) wavelengths, typically expressed in inicrorneters (pm). Optical imagery has the widest range of applications due to its spectral signatures across the visible, NIR, and SWIR regions. It offers rich spectral information of Earth's surface, but its performance can be limited in the presence of clouds, haze, or low-light conditions. The spectral band configurations of widely used optical sensors (U.S. Geological Survey, 2025) (European Space Agency, 2026b) for SWBE are summarized in Table 1. These configurations outline the wavelength ranges across different spectral regions.

1.1.2 Microwave Sensors

Unlike optical sensors, Synthetic Apemire Radar (SAR) sends microwave signals to Earth's surface and records the backscattered response. Particularly, SAR operates in the microwave region and emits electromagnetic pulses toward the Earth's surface. SAR captures back-scattered microwave pulses. It has frequency bands (GHz) and polarization modes (VV, VH, HH, HV) (Hirabayashi et al., 2013). The back-scattered pulse response is governed by surface irregularities and dielectric features. Calm water appears dark due to spectral reflection, whereas rough or vegetated surfaces scatter more energy back to the sensor. SAR has the advantage of being independent of sunlight and largely unaffected by cloud cover. Thus, it can work in all weather, day-night and is useful for flood mapping, continuous water body monitoring, and wetland studies in cloudy conditions, whereas optical sensors struggle in those scenarios. ENVISAT (ENVIronmental SATellite), Sentinel-1, RADARSAT-1,2 (Radar Satellite), TerraSAR-X(Terra Synthetic Aperature Radar-X-band), COSMO-SkyMed (COnstellation of small Satellites for the Mediterranean basin Observation), RISAT(Radar Imaging Satellite), and ALOS PALSAR (Advanced Land Observing Satellite-Phased Array type L-Band Synthetic Aperature Radar) are among the most widely used SAR satellite platforms for SWBE (Mason et al., 2009). The specifications of the microwave sensors (Lillesand, Kiefer and Chipman, 2015)

(European Space Agency, 2026a) are summarized in Table 2. These include operating frequency bands, polarization modes, and spatial resolutions, which play a key role in capturing water body information even under cloudy or low-light conditions.

Table 1: Spectral band configuration of optical sensors, illustrating wavelength ranges and band characteristics.

Satellite/Sensor	Band Name	Wavelength Range (µm)	Spatial Resolution (m)
Landsat 9 OLI	Coastal/Aerosol (B1)	0.430–0.450	30
	Blue (B2)	0.450–0.510	30
	Green (B3)	0.530–0.590	30
	Red (B4)	0.640–0.670	30
	NIR (B5)	0.850–0.880	30
	SWIR1 (B6)	1.570–1.650	30
	SWIR2 (B7)	2.110–2.290	30
	Panchromatic (B8)	0.500–0.680	15
	Cirrus (B9)	1.360–1.390	30
	TIR 1 (B10)	10.60–11.19	100
	TIR 2 (B11)	11.50–12.51	100
Sentinel-2 MSI	Coastal/Aerosol (B1)	0.433	60
	Blue (B2)	0.490	10
	Green (B3)	0.560	10
	Red (B4)	0.665	10
	Red Edge NIR (B5/B6/B7)	0.705–0.775	20
	NIR (B8)	0.842	10
	NIR Narrow (B8A)	0.865	20
	NIR (B9)	0.940	60
	SWIR (B10)	1.375	60
	SWIR (B11)	1.610	20
	SWIR (B12)	2.190	20
SPOT-6	Band (Panchromatic)	0.450–0.745	1.5
	Band 0 (Blue)	0.450–0.520	6
	Band 1 (Green)	0.530–0.590	6
	Band 2 (Red)	0.625–0.695	6
	Band 3 (NIR)	0.760–0.890	6
WorldView-3	Coastal Blue	0.400–0.450	1.24
	Blue	0.450–0.510	1.24
	Green	0.510–0.580	1.24
	Yellow	0.585–0.625	1.24
	Red	0.630–0.690	1.24
	Red Edge	0.705–0.745	1.24
	NIR1	0.770–0.895	1.24
	NIR2	0.860–1.040	1.24
	SWIR1	1.195–1.225	3.70
	SWIR2	1.550–1.590	3.70

Table 2: Frequency band configuration of microwave (SAR) sensors, illustrating their operating wavelengths and band characteristics

Satellite/Sensor	Frequency Band	Wavelength Range (cm)	Polarization Modes	Resolution (m)
Sentinel-1	C-band	3.750–7.500	VV, VH	10
RADARSAT-1	C-band	3.750–7.500	HH	8–100
RADARSAT-2	C-band	3.750–7.500	HH, HV, VV, VH	1–100
ENVISAT	C-band	3.750–7.500	HH, VV	150–1000
ALOS PALSAR-1	L-band	15.00–30.00	HH, HV, VV, VH	10–100
TerraSAR-X	X-band	2.400–3.750	HH, VV	1–18
TanDEM-X	X-band	2.400–3.750	HH, VV	1–18
ALOS PALSAR-2	L-band	15.00–30.00	HH, HV, VV, VH	1–100
RISAT-1A	C-band	3.750–7.500	HH, HV, VV, VH	1–50
COSMO-SkyMed-4	X-band	2.400–3.750	HH, HV, VV, VH	1–100

1.2 Literature Survey Methodology

Literature used in this review was collected from the major scientific databases such as Google Scholar, Springer Link, Scopus, IEEE Xplore, ScienceDirect, and Web of Science. The relevant studies were identified using the following keywords: remote sensing, water body delineation, water body extraction, surface water mapping, water segmentation, satellite imagery, machine learning, deep learning, transformers, and synthetic aperture radar. The review paper prioritized sources from 1996 to 2025, including peer-reviewed journal articles, review studies, and conference proceedings. To ensure quality and relevance, studies directly related to remote sensing-based water body extraction were selected, whereas duplicate records, non-English publications, and studies unrelated to water body extraction were excluded. The selected literature was then grouped into traditional methods, ML methods, DL frameworks, and transformer-based architectures.

1.3 Contributions

Several review articles have addressed the detection of surface water bodies from RSI from various perspectives. Many review articles have examined remote sensing-based surface water body extraction from different perspectives. (Bijeesh and Narasimhamurthy, 2020) surveyed the conventional water extraction methods and discussed spectral-index-based techniques and ML methods for surface water mapping. (Gautam and Singhai, 2024) gave a comprehensive overview of deep learning methods for SWBE. Spectral water indices, SAR-based flood mapping, machine learning classification techniques, and deep learning architecmres for remote sensing applications are some of the specific themes that have been reviewed in the literature (Amitrano et al., 2024) (Purnam, Prasad and Ganasala, 2024) (Maxwell, Warner and Fang, 2018) (Zhu et at., 2017). Although these studies have significantly contributed to the understanding of water body extraction methodologies, most reviews primarily focus on a particular category of methods or a specific application domain. Further, recent studies have focused more on multimodal data fusion, advanced preprocessing techniques, loss functions, benchmark datasets, and evaluation metrics, which are not sufficiently covered in existing surveys. Thus, a new and comprehensive review is required that integrates the progression of water body extraction methods and also investigates the components that improve the model's performance. This review presents a unified and comprehensive analysis of remote sensing-based SWBE techniques. It integrates discussions on remote sensing fundamentals, satellite platforms, optical and SAR imagery, preprocessing techniques, benchmark datasets, loss functions, evaluation metrics, traditional methods, ML methods, DL frameworks, and hybrid CNN-transformer-based architectures. In addition, the review critically evaluates the strengths, limitations, and research trends associated with different methodological paradigms and identifies potential fumre research directions.

The core contributions derived from this research are highlighted below:

i. A comparative discussion of optical and SAR imagery and their roles in improving water body extraction performance under diverse environmental conditions.
ii. This article presents a comprehensive review of SWBE methods ranging from visual interpretation and threshold-based approaches to ML, DL, and hybrid CNN-Transformer frameworks.
iii. A systematic analysis of freely available datasets, preprocessing techniques, loss functions, and evaluation metrics employed in water body extraction studies. All details of these datasets, such as image type, spatial resolution, number of images, and URL links, are given in the Table 6.
iv. A critical assessment of the advantages, limitations, and research challenges associated with existing methodologies.

This review article are structured in the following sections. Section 2 briefly explains the key pre-processing operations, such as atmospheric, radiornetric, and geometric corrections. Section 3 discusses the conventional approaches for SWBE methods, including visual interpretation, thresholding, spectral indices, and classical machine learning algorithms. Section 4 explains deep learning-based approaches, highlighting major architectures. Section 5 presents freely available benchmark datasets and the DL model's parameters used for performance measures in the literature. Section 6 discusses current challenges and limitations, while Section 7 gives potential prospects. Lastly, Section 8 presents a conclusion of the review article.

2 Preprocessing of Remote Sensing Imagery (RSI)

In SWBE tasks, preprocessing is a fundamental step that significantly influences performance because it improves the visual appearance of satellite imagery and makes it suitable for subsequent processing such as feature extraction, classification, segmentation, and other analyses. Raw satellite images typically contain noise, geometric distortions, atmospheric interference, and other factors that can adversely impact the detection of water bodies. Preprocessing steps such as atmospheric correction, radiometric correction, cloud and shadow removal, and image patching enhance spectral contrasts between water and other land-cover classes, which minimize missclassification and improve classification accuracy. RSI, especially at high resolution, is large in size and cannot be processed directly by most neural networks due to memory limitations. Image tiling or patching is a necessary preprocessing operation in DL-based methods for SWBE. The process of dividing a large satellite image into smaller patches or tiles is called image tiling, which makes the data suitable for patch-based learning models (Ramos and Sappa, 2024). This step is especially crucial for high-resolution datasets where the amount of GPU memory is a limiting factor, so that full scene images cannot be used. Therefore, the large image is divided into smaller fixed-size image patches (e.g.,256×256), which may or may not overlap. In overlapping patching, neighboring tiles are extracted with a certain overlap so that edge features are well captured, and objects that are split across tiles are more likely to be captured as one, while in non-overlapping patching, the image is split into fixed-size patches without any overlap between adjacent patches. This method is very efficient and leads to faster processing, but it may introduce boundary artifacts or ignore important features that lie across patch edges (Sherrah, 2016). Overlapping tiles are often preferred in segmentation tasks, where context preservation and boundary accuracy are important. Thus, image patching enables efficient model training and helps deep learning networks learn local water features more effectively, particularly in high-resolution images.

2.1 Atmoshperic Correction

Atmospheric correction is necessary to mitigate the atmospheric effects of the atmospheric particles, including water vapor, aerosols, and gases. Top of atmosphere(TOA) reflectance represents the reflectance values recorded by the satellite sensor at the top of the Earth's atmosphere. It does not accurately represent the Earth's surface characteristics. Therefore, atmospheric correction techniques like the Dark Object Subtraction (DOS) method (Chavez Jr, 1988), Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) (Adler-Golden et al., 1998), and the Second Simulation of a Satellite Signal in the Solar Spectrum (6S) model (Vermote et al., 1997), Moderate-Resolution Atmospheric Transmittance and Radiance code (MODTRAN) (Berk et al., 2009) are applied to TOA reflectance value to derive Bottom of Atmosphere (BOA) or surface reflectance value, which gives a more accurate representation of land and water features. Various algorithms have been developed for different satellite missions. Sentinel-2 Correction Processor (Sen2Cor) is widely used for Sentinel-2 imagery to generate surface reflectance value, whereas Atmospheric Correction for OLI (ACOLITE) is suitable for aquatic and coastal applications (Vanhellemont, 2019) (Louis et al., 2016) (Gascon et at., 2017). Similarly, Land Surface Reflectance Code (LaSRC) and Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) are commonly used to produce surface reflectance values from Landsat imagery (Masek et at., 2006) tVermote et al., 2016). By minimizing atmospheric scanering and absorption effects, atmospheric correction improves the spectral representation of water surfaces.

2.2 Radiometric Correction

The radiometric correction is the next step, after atmospheric correction. It mitigates noise produced by the sensors and changes in brightness values caused by sensor calibration errors. This mistake can be in the form of incorrect pixel values, which may lead to images being overexposed or underexposed. This step adjusts the pixel values to correspond to actual ground reflectance and can be compared between images. The absolute radiometric calibration and relative normalization are applied to remove sensor errors (Campbell and Wynne, 2011). Thus, this process enables pixel values to more precisely represent the surface reflectance properties of an object.

2.3 Geometric Correction

Geometric correction is another important pre-processing step that aligns the satellite image to real-world coordinates through Ground Control Points (GCPs) or Digital Elevation Model(DEM) (Lillesand, Kiefer and Chipman, 2015). This step is required for change detection or multi-source image fusion. Through geometric correction, the satellite image is aligned with a map coordinate system, thereby reducing distortions caused by sensor motion, Earth's rotation, terrain elevation, and view geometry. When DEM is available, ortho-rectification is commonly applied to account for terrain effects and improve positional accuracy. Accurate geometric correction helps align extracted water bodies with reference datasets and reduces boundaries during delineation.

2.4 Image Enhancement and Noise Reduction

The images are also denoised to eliminate noise, such as sensor noise or speckle noise, which degrades the image quality. Image enhancement operations, such as histogram equalization, contrast stretching, and filtering, are

applied to sharpen image quality and reduce noise introduced by the sensor or the environment. Median filters or Gaussian filters are widely used for this purpose. Cloud masking is an additional preprocessing requirement for optical imagery, and tools such as Fmask are commonly used. Noise reduction suppresses unwanted variations in pixel values, improving the discrimination between water and surrounding land-cover classes (Zhu and Woodcock, 2012).
A variety of preprocessing operations are available in open source applications software such as QGIS, Orfeo Toolbox, SNAP (Sentinel Application Platform), and Cloud Compare. ENVI, ArcGIS Pro, MATLAB, and ERDAS Imagine are popular and user-friendly commerical software packages.

3 Surface Water Body Extraction Methods

3.1 Visual Interpretation Method

It is the oldest and most intuitive approach adopted to detect water bodies from RSI. In fact, this method relies on expert knowledge and manual annotation. Therefore, it comprises the manual identification and delineation of water features by domain experts based on perceptual attributes such as shape, size, pattern, color, tone, texmre, and spatial context (Paine and Kiser, 2012a) (Lillesand, Kiefer and Chipman, 2015).

1) Shape and Pattern: Natural water-bodies have irregular shapes while artificial ones (e.g., reservoirs or canals) ofien have geometric boundaries.
2) Color and Tone: Water usually represents uniform dark tones, though this can change with depth, turbidity, and sunlight reflection.
3) Texture: Water surfaces generally appear smooth or homogeneous, in contrast to the rough textures of vegetation or urban areas.
4) Proximity to rivers, terrain depression, or vegetation belts can help infer water presence.

The basic concept of visual interpretation lies in the different spectral and spatial behaviors of surface water bodies. The experts use False Colour Composite (FCC) images (e.g., including the NIR band) to refine the visibility of important features for water bodies. Dark tones, smooth textures, irregular or geometric shapes, and contextual surroundings all aid in making accurate judgments (Smith, 1997). This approach produces high accuracy at small scales, making it ideal for localized analysis and ground-truth validation.

3.2 Threshold-based Methods

Thresholding is widely adopted and the simplest approach for SWBE. It involves classifying pixels as water or non-water depending upon a specific reflectance or index value. It is a pixel-based classification method in which a specific value (or range of values) is used as a cutoff to separate the target and non-target classes. An image, $f(x, y)$, a pixel at a given location(x, y) in the plane is referred to as either a water pixel or a non-water pixel if $f(x, y)$>T. This threshold is typically applied to individual spectral bands or to spectral indices that enhance water features. The threshold value can be manually selected, determined statistically, or optimized using techniques like Otsu's method (Sekertekin, 2021) or histogram analysis. Manually selected thresholds may not generalize well across different scenes, leading to misclassification in areas with complex land-water interfaces, turbid water, or urban shadows.
There are mainly two types of thresholding methods: single-band thresholding and spectral index-based thresholding (Bijeesh and Narasimhamurthy, 2020).

- Single Band Thresholding Method: The single band thresholding technique is one of the most fundamental techniques for identifying surface water bodies in RSI. In this method, the pixel values of a particular spectral band, usually the NIR or SWIR bands, are directly subjected to a threshold. Water absorbs radiation in these bands with high strength, and hence the reflectance values are very low, which is clearly contrasted to the land feamres like vegetation, soil, or built-up areas that are highly reflective in nature (McFEETERS, 1996). The most important feature of the single-band technique is selecting the best spectral band for SWBE. This selection is based on the properties of the sensor, the nature of the landscape, and the spectral signature of water. After selecting the optimal spectral band, a thresholding technique is applied to isolate water pixels from other land cover types.

- Spectral Indices Thresholding Methods: Spectral index-based thresholding is based on the threshold value of computed spectral indices, such as the normalized difference water index (NDWI), which is a mathematical combination of several spectral bands used to strengthen water features and suppress background noise caused by vegetation, built-up areas, or soil. These spectral indices leverage the distinct spectral signature of water: strong absorption in the NIR and SWIR bands, while high reflectance in the Green and Blue bands. These methods offer the most basic computational approach for detecting surface water bodies. Over the years, researchers have designed many indices to handle different imaging conditions and land cover complexities. The NDWI developed by (McFEETERS, 1996) is the basic index technique, which is calculated as the normalized difference of the reflectance in the Green and NIR bands.

$$NDWI = \frac{(B_3 - B_5)}{(B_3 + B_5)}$$

where B_3 and B_5 indicate the intensity values in the green and near-infrared bands of Landsat 9 OLI, respectively. This NDWI works successfully on open water features in most RSI but may be prone to noise produced by built-up surfaces, shadows, and vegetation. To overcome this drawback, (Xu, 2006) proposed the Modified NDWI(MNDWI) that substitutes the NIR band with the SWIR band, which enhances the accuracy of water detection in urban and mixed land cover regions considerably.

$$MNDWI = \frac{(B_3 - B_6)}{(B_3 + B_6)}$$

where B_3 and B_6 denote the intensity values in the green and shortwave infrared bands of Landsat.
Other spectral indices have been developed to address specific challenges. The Automated Water Extraction Index (AWEI) is another important index that was developed by (Feyisa et at., 2014). It is quite useful in minimizing false positive errors due to shadows and dark surfaces. It has two variants: AWEI-sh for shadow-prone areas and AWEI-nsh for non-shadowed areas. Both versions of the index incorporate multiple bands in weighted formulations to enhance water features and reduce background noise.

$$AWEI_{sh} = B_2 + 2.5B_3 - 1.5(B_5 + B_6) - 0.25B_7$$

$$\mathrm{AWEI_{nsh}} = 4(B_3 - B_6) - (0.25B_5 + 2.75B_7)$$

where B_2,B_3, B_5 B_6 and B_7 represent the intensity values in their respective spectral bands of Landsat 9 OLI .
In flood plains where water is turbid or contains has high sediment loads, Normalized Difference Turbidity Index (NDTI) and Normalized Difference Pond Index(NDPI), which use SWIR-1 and Green bands are employed to identify sediment-laden water (Lacaux et al., 2007).

$$\mathrm{NDTI} = \frac{(B_2 - B_1)}{(B_2 + B_1)}$$

where B_2 and B_1 denote the reflectance values of the red and green bands of SPOT-5 satellite, respectively.

$$\mathrm{NDPI} = \frac{(B_4 - B_1)}{(B_4 + B_1)}$$

The Normalized Difference Pond Index (NDPI) enhances the detection of small and shallow water bodies such as ponds and wetlands by contrasting green and shortwave infrared (SWIR) reflectance, where B_4 and B_1 denote the reflectance values of shortwave infrared(SWIR) and Green bands of SPOT-5 satellite (Lillesand, Kiefer and Chipman, 2015), respectively.

In order to improve water detection, multivariate indices like the Tasseled Cap Wetness(TCW) (Crist and Cicone, 1984) index have been used. It is an index based on a linear transformation of multiple spectral bands and is useful in detecting moisture-related features such as water or wetlands, particularly in the presence of vegetation cover or in mixed-pixels. It is particularly sensitive to water, wet soil, and vegetation moismre.

$$\mathrm{TCW} = 0.1509B_1 + 0.1973B_2 + 0.3279B_3 + 0.3406B_4 - 0.7112B_5 - 0.4572B_7$$

where B_1, B_2, B_3, B_4, B_5, and B_7 denote the intensity values in the corresponding bands of the Landsat TM satellite.

The Water Index (WI) is another spectral method that enhances water body detection by suppressing noise from vegetation and built-up areas. It was formulated by the combination of the TCW and NDWI to extract the shoreline changes on Landsat TM and ETM+ Imagery (Ouma and Tateishi, 2006).

$$\mathrm{WI} = \alpha \times \mathrm{NDWI} + \beta \times \mathrm{TCW}$$

Where α and β are are weighting coefficients that balance the contribution of NDWI and TCW respectively Typically equal weighting ($\alpha = \beta = 1$) or empirically derived coefficients are used depending on the sensor and study area. The Water Index (WI) particularly valuable in distinguishing water from surrounding land surfaces, which often confuse other spectral indices.

Another index, called NDMI, was proposed by (Gao, 1996). The NDMI, originally designed for vegetation moisture content, has been adapted for detecting wet or shallow water surfaces, though it may require complementary indices to distinguish from moist soil.

$$\mathrm{NDMI} = \frac{(\mathrm{NIR} - \mathrm{SWIR})}{(\mathrm{NIR} + \mathrm{SWIR})}$$

NIR and SWIR denote near-infrared and shortwave infrared reflectance values of the bands, respectively. In a similar way, there is an alternative spectral index called the Normalized Difference Infrared Index (NDII). NDII is specially developed to detect the moisture content in vegetation and soil. NDMI and NDII both indices use the reflectance differences between the NIR and SWIR bands. Water and moisture-rich features absorb SWIR and reflect NIR in RSI, leading to a higher value of NDMI in these regions. However, dry soil or vegetation produces a negative value of NDMI. This characteristic makes NDMI especially useful in detecting the presence of water in a mixed or vegetated environment like flood plains, paddy fields, and marshes.

The most critical aspect of implementing spectral indices is that an appropriate threshold value must be selected to identify the water and non-water pixels. This value can be manually adjusted, usually a range of 0 to 0.5 in the case of NDWI, based on visual inspection or empirical analysis. On the other hand, automated thresholding methods like Ostu's method or K-means clustering are adopted to dynamically calculate the threshold value based on the statistics of index values in a given image scene.
In a similar manner, (F. Zhang et al., 2018a) put forward a modified histogram bimodal method (MHBM) which automatically gives the threshold value. This MHBM overcomes the difficulties associated with complex surfaces, image noise, and the imperfect twin peaks ofien encountered in histogram bimodal methods. First of all, RSI undergoes a Rayleigh scanering correction to obtain the reflectance value, and is then used to calculate water indices like NDWI, MNDWI, and others. After that, the initial threshold value is calculated to provide a rough mask of water identification. To ensure well-defined twin peaks in histograms, the rough water masks are expanded by 1.5 times to create a work area buffer. Instead of relying on twin peaks, the method seeks minimum values within the identified threshold range to determine precise thresholds for SWBE. The main contributions of MHBM include its simplicity, low data requirement, and high accuracy, which make it suitable for mass processing of RSI.
The main drawbacks of these methods are their sensitivity to environmental conditions, including atmospheric noise, sensor variability, and seasonal or regional differences in surface reflectance.

3.3 Machine Learning Methods

ML Methods: ML algorithms learn patterns from the labeled data and train well so that they can adapt more to complex and heterogeneous land cover environments. These algorithms have the widest applications, especially where the spectral ambiguity of water bodies is present due to urban shadows, dark soil, and vegetation. These methods rely heavily on feature engineering, where spectral, spatial, and texture descriptors are crafied from raw image bands. Standard feature sets for SWBE comprise raw reflectance values, spectral indices like the NDWI, MNDWI, and NDVI, GLCM-based textural feamres, shape descriptors for connected components, and topographic attributes from DEMs. The choice and combination of these features often determine the success of an ML-based water mapping model. The key ML models explored in the literature are summarized in the Table 3.

3.3.1 Supervised ML Methods

Supervised ML requires labeled datasets for training, where each input (e.g., satellite image pixels) is associated with a known output label (e.g., water or non-water). These methods have gained popularity since the early 2000s, with increasing improvements in classification accuracy. (Frazier and Page, 2000) applied Maximum Likelihood Classifier (MLC) and single band density slicing to detect boundaries of open water bodies in riverine floodplains using Landsat 5TM data. The authors made a comparison of the performance of single mid-infrared band 5 density slicing and a multi-spectral 6-band MLC. The single band 5 achieved an Overall Accuracy (OA) of 96.9%, a Producer's Accuracy (PA) of 81.7%, and a User's Accuracy tUA) of 64.5%.
(Huang, Davis and Townshend, 2002) employed SVM on Landsat TM images to identify water class and compared its performance with other classifiers like MLC, Decision Tree Classifier (DTC), and neural network classifier (NNC). The impact of different kernel configurations on the performance of SVM and input variables, training data on these classifiers, is also analyzed. For 20% training data and 7 variables, SVM produced a mean overall accuracy of 75.62%. (Sarp and Ozcelik, 2017) used SVM and spectral indices like NDWI, MNDWI, and AWEI to study spatio-temporal changes in Lake Burdur, Turkey. Landsat TM and ETM+ imagery of the lake from 1987 to 2011 was acquired for analysis. Spectral and spatial performance of the classifiers was evaluated using Pearson's correlation coefficient (r), the Structural Similarity Index Measure (SSIM), and Root Mean Square Error (RMSE). The SVM, followed by spectral indices methods such as NDWI, produced the superior results in terms of spectral and spatial quality.
(Li and Narayanan, 2003) suggested a shape-based approach for change mapping of a lake using time series multi-spectral images. The study examined seasonal and interannual variations of lakes by utilizing a time series of Landsat Multi-Spectral Scanner System(MSS) (1981-1987) and TM (1992-1997) images from the Nebraska Sand Hills region. The authors integrated an SVM classifier to generate a water mask, connected-component labeling for object recognition, parametric contour tracing, and a piecewise linear polygonal approximation (PLPA) algorithm to simplify lake boundaries. The SVM classifier's OA on MSS and TM images is 91.5% and 93.6%, respectively.

(Fu, Wang and Li, 2008) developed a SWBE method using the DTC algorithm. The authors analyzed the spectral attributes of water bodies in different TM bands and developed the relational formula (band2+band3) > (band4+band5) to isolate water bodies from other features. The study showed that the model attained a PA of 81%, a UA of 94%, and a kappa coefficient of 92%. However, the proposed model fails to detect small water bodies and misclassifies boundary regions.
(Mishra and Prasad, 2015) used the perceptron model to classify pixels of an image: water and non-water. Three feature vectors: Spectral band relationships, low reflectance radiance of water in the SWIR band, and MNDWI are selected to train the perceptron model and improve pixel classification. The experiment is conducted on the Landsat ETM images of Hyderabad city, India. The OA of the experiment was 94.4%. (Malinowski et al., 2015) addressed the key critical challenge of mapping localized flooding in intricate river floodplains where patches of herbaceous vegetation are mixed with open water surface. The DTC is employed to detect and delineate the flooded area from WorldView-2 imagery. The study highlights that single-date WorldView-2 image data, despite lacking SWIR bands, can reliably delineate open water and mixed water-vegetation patches, especially when combined with OBIA and elevation data, but mapping inundation beneath closed vegetation canopies remains problematic.

(Rokni et al., 2015) applied the image fusion technique at the pixel level and then classified the fused images. The multi-temporal Landsat ETM+panchromatic and Landsat TM multi-spectral images of Lake Urmia were fused in this study. The authors used Artificial Neural Network (ANN), SVM, and MLC to detect and map the change detection. The research study revealed that the surface area of Lake Urmia reduced by one-third between 2000 and 2010.
(Trochim et al., 2016) mapped the location of water tracks in the foothills of the Brooks range in Alaska using an RF model. The field data, WorldView-2 multi-spectral imagery, and DEMs were used to extract spectral, textural, and topographic features. These features were fed to an RF model. The study showed that WorldView-2 imagery at 1.84 m resolution can capture the curvilinear panems of water tracks and provide meaningful insights into their geomorphic and ecological associations, despite errors of omission and class overlap. (Paul, Tripathi and Dutta, 2018) conducted their study across three sites in India, utilizing multi-sensor data such as LISS III, AWIFS, and Landsat 8 OLI for feature representation and cross-validation. The authors analyzed the performance of ANN, SVM, K-NN, MLC, and RF with the traditional spectral method, NDWI. Supervised classifiers demonstrate superior and more consistent performance compared to NDWI across all study areas and sensor types, even with only a few labeled samples.
(Nagaraj et al., 2023) addressed the issue of quantifying the volume of surface water bodies using RSI. They estimated the volume and surface water extent of Lake Victoria by RSI and bathymetry data. The predicted water maps are obtained using various ML algorithms like Gaussian Naive Bayes (GNB), DT, RF, XGBoost, Light Gradient Boosting Machine (LightGBM), and CatBoost. The results reveal that LigbtGBM is the best ML algorithm for estimating the surface extent and volume of the lake. The LightGBM achieved accuracy, precision, recall, and mean intersection of unit (mIoU) of 97.5%, 98.0%, 96.9%, and 95.0%, respectively.

3.3.2 Unsupervised ML Methods

Unsupervised ML techniques are particularly useful when labeled water body datasets are unavailable. These approaches aim to discover inherent data structures, such as clusters that belong to target and non-target regions. It creates clusters and classifies pixels based on spectral and spatial similarities.
(Gao, Birkett and Lettenmaier, 2012) utilized a k-means clustering approach to compute the surface water area of reservoirs from MODIS 250m vegetation index images. The authors first applied a normal NDVI approach to get a temporary water map and then applied the k-means clustering approach to get the refined map. K-means thus provided a non-parametric, globally consistent classification of reservoir surface areas, which were then combined with altimetry-derived elevations to develop elevation-area relationships and storage time series. (Yousefi et al., 2018) employed a cluster-based method to isolate surface water bodies from RSI. Principal Component Analysis (PCA) and the k-means clustering were utilized to derive important features from the satellite data. The cluster-based approach is effective for open-water detection, but its misclassification rate is high for urban water bodies. A limitation of k-means clustering is the lack of semantic understanding.
A new automatic algorithm for mapping urban surface water bodies (AUWM) was developed in 2022 by (Gasparovic and Singh, 2023). The MNDWI, which is calculated through pansharpened band and the k-means clustering, is the foundation of the AUWM. The study discovered that four classes are the ideal number for k-means clusters in AUWM. The AUWM attained OA, a kappa coefficient, and a precision of 0.997, 0.830, and 0.998, respectively.
The ISODATA (Iterative Self-Organizing Data Analysis Technique) algorithm is a well-recognized clustering method that extends basic k-means by dynamically partitioning and combining clusters based on variance and inter-centroid distance. Originally introduced by (Ball and Hall, 1965) (Bezdek, 1980), ISODATA has proven useful in SWBE, where it groups pixels that share similar spectral or backscatter properties.
(Sivanpillai and Miller, 2010) evaluated the performance of the ISODATA classification technique on higher resolution ASTER imagery and Landsat TM imagery. The study shows that the ASTER imagery significantly outperformed Landsat TM imagery in detecting small, clear, and green water bodies. The ISODATA classification achieved PA and UA of 94% and 100%, respectively, for large-scale water-bodies in Landsat imagery, while on ASTER imagery, it achieved PA and UA of 97% and 100%, respectively. (Hong et al., 2015) suggested an effective approach for extracting flooded water areas from RADARSAT-1 SAR imagery. It mainly relies on the thresholding method, which removes noise by using SAR amplitude and terrain data. The authors utilized the ISODATA algorithm to generate a land cover map, which was then used to determine appropriate cut-off values for classifying water regions. The method still demands manual identification of the water cluster and often requires post-processing (e.g., morphological filtering or edge-based refinement) to achieve clean water segmentation.
A fuzzy classification employed fuzzy sets for the classification purpose. Fuzzy logic classifies data points into another fuzzy set based on the degree of membership of the set to which they belong. Fuzzy C-Means(FCM) clustering, first introduced by (Bezdek 2013), has the widest applications in SWBE methods due to its ability to assign partial memberships to multiple classes, which makes it effective for mixed pixels and complex land—water boundaries. (Zhang and Foody, 1998) present and evaluate a full fuzzy strategy for classifying a suburban land cover map from RSI. The authors utilized the FCM clustering algorithm to generate fuzzy membership values for each pixel and compared the results with traditional hard classification approaches. Experiments were conducted using SPOT HRV and Landsat TM imagery over a suburban area in Edinburgh, UK. An accuracy of 92.8% and a kappa coefficient of 90.0% are achieved using SPOT HRV data, while an accuracy of 93.2% and a kappa coefficient of 90.1% are achieved using Landsat TM data.
(Nedeljkovic, 2004) applied a fuzzy logic approach to classify SPOT imagery into different land cover (LC) classes. A priori knowledge regarding spectral information on specific LC classes was used to classify the images using fuzzy logic. The SPOT imagery was initially classified through the MLC, and the resulting supervised classification outputs were subsequently used to derive fuzzy membership functions for fuzzy land cover classification.

(Pierdicca et al., 2008) suggested a flooded area identification method based on fuzzy logic concepts. The authors utilized SAR images along with DEM data to map flooded areas. Two scattering effects, the specular reflection from the open water surface and the double bounce in flooded areas and forest regions, are incorporated in the experiment to differentiate land cover types. The different fuzzy sets, such as dark pixels and high difference, are employed in the study. The weighted mean of fuzzy membership values is calculated.
In the same series, (Singh and Singh, 2017) developed a novel hybrid Kohonen fuzzy c-means-σ(HKFCM-σ) classifier. It includes descriptors of both a fuzzy system and a neural network. The Jammu and Kashmir region of India is selected as the study area. The authors extracted spectral features such as NDVI, NDWI, NDBI, and PCA-1 from the RSI to train the fuzzy classifier. The proposed classifier was tested on Landsat 8 OLI images to identify the flooding regions. The classified images were subsequently analyzed using a post-classification comparison approach to produce a change map, which highlighted the flooded regions.

4 Deep Learning Approaches

Unlike ML approaches, which work on user-defined features, the DL approaches, particularly CNNs and their variants, can accurately segment water bodies from diverse RSI datasets without requiring hand-crafted features. DL encompasses a wide range of neural network architectures. Among them, CNN-based models directly extract and combine spectral-spatial features. It extracts hierarchical descriptors such as shapes, texmres, and edges while learning local spatial patterns including shoreline boundaries and the spectral variations between water and vegetation or soil. FCNs have become one of the milestones for SWBE. Other variants, such as U-Net, have also been developed that enhance the performance by including skip connections, which retain spatial information that is lost during down-sampling layers. Later, there was another FCN architecture, called DeepLabV3+, which added atrous(dilated) convolutions, spatial pyramid pooling to handle multi-scale features, enhanced detection of water bodies of different sizes and time-varying spectral appearances. In recent work, transformer-based architectures have been proposed. In contrast to DL models, transformer models can capture long-range dependencies across the whole image scene. These architectures are adaptable to different spatial resolutions, sensors, and geographical conditions. Figure 5 shows a simple flowchart of the entire process of surface water body detection from RSI. It begins with collecting satellite images and follows with preprocessing steps, which were already discussed in earlier sections. The flowchart then proceeds to the segmentation stage, where different DL architectures are applied to isolate water bodies from other land cover types. This diagram helps readers to understand the process step by step. It also connects all the important parts of the workflow discussed throughout the review. CNNs act as fundamental components in a wide range of DL models. The key DL models explored in the literature are summarized in the Table 4. Although they cannot be strictly categorized, they are typically classified into three broad types according to their layers of convolutional operations, pooling mechanisms, and architectural design, as illustrated in the Figure 6. In general, the merits and demerits of each SWBE method are summarized in Table 5.

4.1 CNN Based Methods

The major blocks of any CNN model are shown in the Figure 7. The CNNs offer automated feature extraction capabilities and reduce reliance on handcrafted indices.

(Paisitkriangkrai et at., 2016) presented, for the first time, a CNN-based approach for pixel-level classification of aerial and satellite imagery. The framework comprises 4 convolutional layers, 2 frilly connected layers, and a 5-way softmax operation. Handcrafted features such as NDVI, saturation, entropy, and normalized digital surface model (nDSM) were extracted. The handcrafted features, deep learning features, and Conditional Random Field (CRF) post-processing were combined to refine the object boundaries. The proposed framework achieved 88 % overall accuracy on the ISPRS Semantic Labeling challenge dataset. A multi-resolution approach improves accuracy but also increases computational cost.(Yu et at., 2017) designed a CNN model for surface water body identification. This model is a combination of a CAN and a logistic regression(LR) classifier. CAN model, with they abiliy to learn hierarchical deep features, offers a more automated and powerful alternative by directly learning spectral and spatial patterns from raw image patches. This proposed model is superior to SVM and ANN and offers a more generalized solution than spectral indices-based approaches. However, the segmented image patches introduced some redundancy.
(Kernker, Salvaggio and Kanan, 2018) suggested a CNN-based model for the semantic segmentation of MSS images. The authors developed their own synthetic dataset with 18 object classes. The RIT-18 dataset is used to initialize a CNN model for object recognition. Two advanced semantic segmentation architectures, SharpMask and RefineNet, were adapted for Multi-Spectral Imagery (MSI) and compared with traditional classifiers such as kNN, SVM, and Multi-layer Perceptron (MLP), MICA, and SCAE. RefineNet-Sim model has achieved the highest mean class accuracy of 59.8%, considerably outperforming other models such as KNN, SVM, and MLP. The models struggled with small object classes (e.g., black panel, pond, low vegetation). The segmentation maps still suffered from salt-and-pepper label noise and boundary misclassifications. (Yang et al., 2020) introduced a mask R-CNN-based deep learning model to identify the surface water bodies. The authors concentrated on the limitations of RSI, such as low resolution, irregular shapes, and complicated backgrounds. To tackle these issues, a Mask R-CNN is developed that adopts two backbone networks, ResNet-50 and ResNet-101. A dataset was constructed using RGB imagery from Google Earth, NWPU VHR-10, and Gaofen-2, which was then used to train the model. The results indicated that the ResNet-fi0 surpassed ResNet-101, achieving an average IoU of 90.4% for regular-shaped water bodies and 84.5% for irregular or fragmented ones. Both models are unable to detect multiple water bodies from the image sample.

Table 3: Summary of ML approaches for SWBE, including satellites, study area, results, and findings

ML Models	Dataset	Study Area	Results	Findings	Aulhor
Ensemble Learning — Adaboost algorithm	Landsat ETM+ images	Huairou reservoirs and Bengcuo Lake, Beijing, China	Qualitative result is analyzed.	Reduces mountain shadow confusion, built-up land, and accumulated snow but requires more processing time than single-feature approaches.	*(Shen and Li, 2010)*
FCM, SVM	Landsat-7, 8	Tamil Nadu, India	OA: 94%	Highly sensitive to noise.	*(Vignesh and Thyagharajan, 2017)*
MLC, Visual interpretation	Landsat-5, SPOT-5, FCC image	Pearl River Delta (PRD)	Change detection is studied.	The surface area of the PRD has been notably reduced due to human activities and urbanization.	*(Lian and Jianfei, 2011)*
DTC	Landsat-OLI	Han River Basin, Gangwon, Korea	OA:99.1%, Kappa: 98.3%	Unable to detect narrow rivers and mixed pixels at boundaries.	*(Acharya et al., 2016)*
SVM, DTC, RF, TB, ELM, & LORSAL	GeoEye-1, WorldView-2	Wuhan, Shenzhen, China	OA:96.2%, Kappa:95.4%	This approach is highly suitable for complex urban environments.	*(Huang et al., 2015)*
Thresholding, DTC, KNN, RF, SVM	SAR and Sentinel-2 MSI images	Cape Winelands, South Africa	OA:91.7%, Kappa: 82.0%	The integration of SAR and Sentinel-2 images improves identification of water from other objects in complex environments.	*(Bangira et al., 2019)*
Naive Bayes, RF, SVM, XGBoost	LISS-III	Veeranam Lake, Tamil Nadu, India	XGB classifier is best among other classifiers.	The XGB ML model is unable to detect small water bodies and misclassifies land areas.	*(Nagaraj and Kumar, 2021)*
NDWI, K-means clustering	Geofen-1 satellite images	Suizhong County, Hangzhou City, China	Developed new spatial index called Pixel Region Index (PRI).	The MFWE method mitigates issues such as distinguishing water from built-up areas and improving the accuracy of water body borders.	*(Y. Zhang et al., 2018)*
FCM	MODIS	Karkheh River, *Iran*	Various indices like SAVI, NDMI, and LAI have been calculated.	Effective for watershed identification, flood management, and soil conservation.	*(Choubin et al., 2017)*

(Zhang, Li and Hua, 2022) put foru ard a CNN model that uses the concepts of multi-scale residuals and squeeze and excitation(SE) mention mechanism. Encoder captures contextual details at multiple scales using a chain of 3×3 convolution kernels. It also incorporates residual connections with 1× 1 convolution. The SE-attention module first applies global average pooling to each channel to generate global compressed feature. Improved skip connections that pass the lost information during the downsainpling process to the decoder section. The model was evaluated on Landsat-8 imagery and its performance was validated against the Global Surface Water Dataset. A novel CNN architecture named "WaterNet" was proposed by (Farooq and Manocha, 2025). Dense layers(dense down units and dense up units) have been introduced in the encoder and decoder sections to reduce the total model's parameters and alleviate computational load. To de-noise the images, two residual refinement modules are incorporated in the WaterNet. By integrating dense layers, asymmetric convolutions, and residual refinement modules, WaterNet achieves high accuracy (97%) and significantly reduces computational costs. Experiments were conducted on three publicly available datasets: CIFAR- 10, Aquasat, and COCO-10. WaterNet demonstrated superior performance compared to various approaches, including U-Net, DeepU-Net, SegNet, DeepwaterMap, NDWI, and Multi-layered Feature Fusion Approach (MFFA).

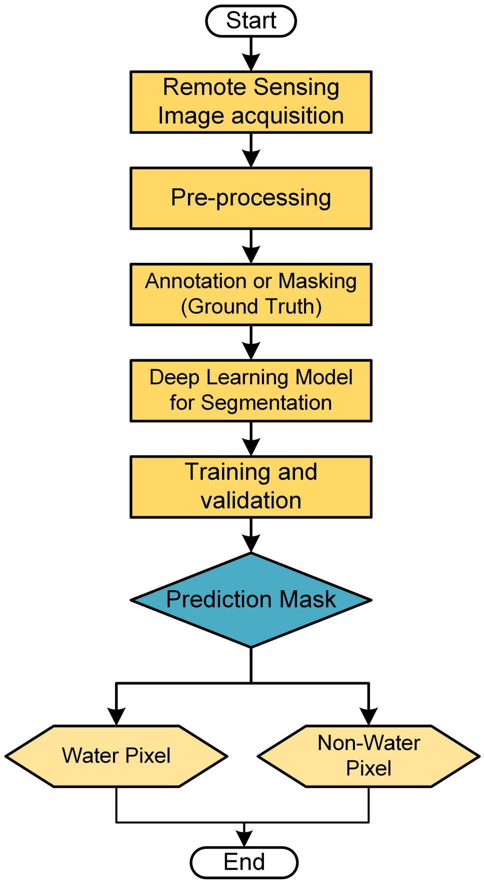

Figure 5: Flowchart showing the overall process of water body extraction from satellite image acquisition to prediction of a water map

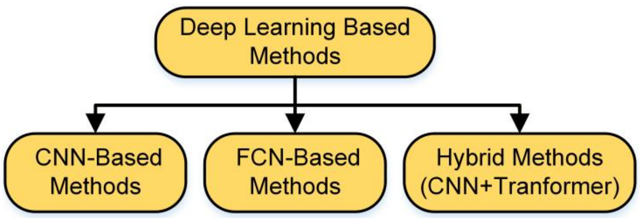

Figure 6: Taxonomy of deep learning methods based on three major categories

4.2 Fully Convolutional Network (FCN) based Methods

Traditional CAN models for SWBE require input images of the same size because the frilly connected layers at the end of the neuvork enforce this requirement. To handle this limitation, researchers often split large images into small patches and classify each patch as water or non-water. However, this patch-based approach lost global spatial context and ofien produced inaccurate water boundaries. Furthermore, FCN proposed by (Long, Shelhamer and Darrell, 2015) removed the fiilly connected layers and incorporated only convolution and upsarnpling layers. This modification enabled FCN to work with any input-size image and generate dense, pixel-level segmentation maps aligned with the original input image. (Ronneberger, Fischer and Brox, 2015) introduced U-Net, a landmark architecture for semantic segmentation that combines an encoder—decoder structure by skip connections, which retain fine spatial details. The basic architecture of the FCN-based model is shown in Figure 8.

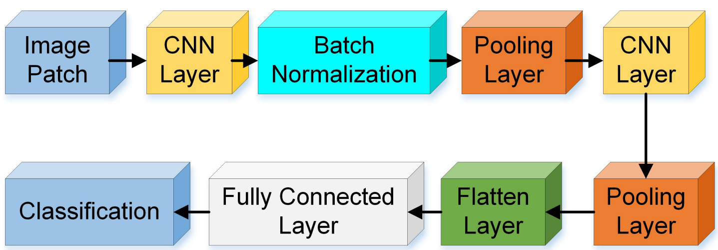

Figure 7: Basic CNN architecture showing convolutional, batch normalization, pooling, flatten, and fully connected layers

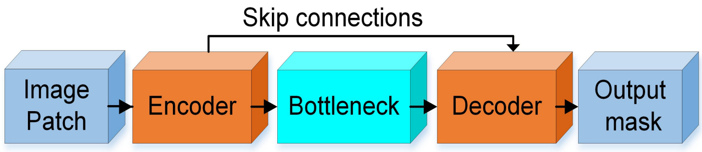

Figure 8: Architecture of basic fully convolutional network (FCN) for semantic segmentation

(Feng et at., 2018) recommended a hybrid framework, which is a combination of an improved version of U-Net called Deep U-Net, a fully connected CRF, and Regional Restriction (RR) based on super-pixels. A hybrid framework had 33 convolution layers. The ELU activation function was utilized for the training. The U-Net architecture learns hierarchical spatial features, while the superpixel-based CRF refines boundaries and reduces noise, leading to more precise delineation of surface water bodies, especially in shadowed and mixed land-cover areas.

(Z. Li et at., 2019) focused on the limitations of single-scale training that lead to reverse errors and imprecise details. The authors proposed a novel deep learning scheme that involves a hierarchical multiscale splicing method to enrich detailed features, an optimized DeepLabV3+ model, and a fiilly connected (CRF). ResNet-50 and Atrous Spatial Pyramid Pooling (ASPP) are utilized in the encoder section of optimized DeepLabV3+. A fully connected CRF is applied as a postprocessing module to refine water boundary details. Thus, CRF sharpens boundaries and removes the noise. This research study integrated multiscale feature processing, an optimized DeepLabV3+ model, and CRF postprocessing to achieve high precision and high classification accuracy on Geofen-1 RSI.

(Yuan et at., 2021) developed a multi-channel network for water body detection, referred to as MC-WBDN. The proposed model integrates three innovative modules, namely: the Multichannel Fusion model, which combines RGB, NIR, and SWIR bands without upsampling, an Enhanced Atroiis Spatial Pyramid Pooling(EASPP) module designed to capture multi-scale contextual feature details, and Space to Depth(S2D)/ Depth to Space(D2S) transformations, which substitute conventional pooling and upsampling layers for better feature retention. The ResNet-34 model acts as an encoder of MC-WBDN. The MC-'ISDN model is optimized using Dice loss, Jaccard loss, and Binary Cross-Entropy loss, and obtains a mean Intersection of Union (mIoU) of 74.42%. The findings reveal that leveraging multispectral information substantially improves the detection of fragmented and intricate water bodies.

(Filatov and Yar, 2022) utilized a U-Net-based model for monitoring envñomnental changes like deforestation and areas affected by flooding. The proposed neuvork has 18 CAN layers. MSS images from the DeepGlobe Challenge and Sentinel-2 images were collected and employed for model training. The OA is 82.55% and 82.92% for the forest region and water region, respectively. The study focuses on reliable labeling and highlights the U-Net's suitability for fast and precise semantic segmentation.

(Ge, Xie and Meng, 2022) developed an improved U-Net architecture and applied it to a China Lakes and reservoñs (LaR) dataset. The Lakes and Reservoirs (LaR) dataset was developed from Geofen-1 four-band imagery. The basic U-Net architecture was modified by expanding the number of convolutional layers in both the encoder and decoder, while incorporating local skip connections to reuse features from deep layers and capture more comprehensive information. The performance of an improved U-Net model is compared with other models like U-Net, U-Net+-r, FastFCN, NDV9, and DeepLabv3+. (Kothari et al., 2026) investigated the adversarial robustness of DL models for SVvTtE using SAR data. This paper primarily contributes to understanding the adversarial robustness of the models, specifically the U-Net, for inland water body identification from SAR images by simulating manual annotation errors through controlled morphological operations (erosion and dilation) on training data. The U-Net model exhibits significant robustness, tolerating up to 17% corruption in labels before its performance notably degrades. However, the models struggle to frilly delineate narrow tributaries.

(Han et al., 2024a) presented an innovative attention fusion U-Net architecture (AFUNet). ResNet-18 serves as the backbone of AFUNet's encoder and captures relevant features. The decoder recovers the original input image resolution while maintaining semantic information. Attention mechanisms(Spatial attention and Channel attention) are inserted within skip pathways beuveen the encoder and decoder. Spatial mention module(SAM) extracts low-level spatial features in the initial Evo layers, while Channel attention module(CAM) captures high-level features in the subsequent three layers to counterbalance for semantic information loss during down-sampling and guide the reconstruction process in the decoder. The authors integrated two more blocks, namely the multiscale convolutional pooling block(MCPB) and the Attentional upsampler (AU). A novel Active Contour(AC) loss function is also designed to optimize the target contours. Experiments were performed on ALOS PALSAR and Sentinel-1 SAR datasets. Model's performance is compared with other DL models such as U-Net, PSPNet, SegNet, AUNet, DeepLabV3+, ResUNet, and DAUNet. On the ALOS PALSAR dataset, AFUNet achieved an overall accuracy of 96.34%.

4.3 Hybrid Deep Learning Methods (CNN-Transformer)

Transformer-based models, originally proposed by (Vaswani et al., 2017) for Natural Language Processing (NLP) tasks, have found the widest applications in image segmentation, because they can team long-range dependencies and global contextual relationships without relying on convolutional operations. The CNN-based model fails to learn global contextual information but is more effective in learning local feamres, whereas the Transformer is more efficient in learning global contextual information. In this way, the network combines the local feature-learning capabilities of CNNs with the ability of transformers to team global contextual information. The adaptation of Vision Transformer (ViT) (Dosovitskiy et al., 2020) for water mapping began in the early 2020s, with Swin Transformer (Liu et al., 2021), with its hierarchical architecture and shifted window mechanism, was first explored for SWBE. Another notable contribution came from SegFormer by (Xie et al., 2021), a lightweight

transformer encoder with a simple decoder. Noise Cancelling Transformer (NT-NET) is an end-to-end semantic segmentation network that was designed by (Zhong et al., 2022). To address challenges such as blurred boundaries, cloud shadows, terrain, and seasonal variations, NT-NET includes two novel modules: the Interference Attenuation Module(IMA) and Multilevel Transformer(MT). IAM is developed to tackle the issue of over-segmentation caused by other land objects, whereas MT captures global boundary information across multiple semantic levels to produce fine, continuous lake boundaries. Multiple features were extracted using residual CNN blocks with both average and max pooling layers.
(Zhang et at., 2023) developed a transformer-based model, Multi-scale Fusion SegFormer (MF-SegFormer), which is a combination of the SegNet model and the Transformer architecture. The main components of the architecmre are a feature fusion module and ASPP, which effectively isolate surface water areas. The feamre fusion module fuses the feamres at different scales so that it can capture richer semantic information and enhance edge information. **ASPP** improves the identification of fine-scale water features through multiscale contextual aggregation. The major limitations of MF-SegFormer are high computational cost and huge memory requirements.
(Wang et al., 2024) adopted the concept of quatemion algebra in convolution operations to detect surface water from RGB satellite imagery. The authors developed a U-Net architecture based on quaternion transformers called the QTU-Net model. The model has three main modules, namely the quaternion initialization module (QIM), the quatemion convolution block (QCon block), and the rnultiscale aggregation attention module (MSSA). The QIM encodes an RGB three-channel image in quatemion data. The QCon block performs a convolution operation on quatemion data generated by the QIM. The MSSA uses a multiscale downsampling method to aggregate global information and compare it with local details to obtain local similarity. The QTU-Net is trained on three publicly available benchmark datasets, such as WHDLD, GLH-Water Dataset, and Water Bodies Dataset.

(Wang et at., 2025) implemented a masked hybrid network(MHNet), which integrates CNN and ViT. The MHNet proposes an MCFF (multi-channel feature fusion) module to synthesize masked feature maps across diiTerent scales, reducing redundancy and improving generalization. Restorrner blocks comprising Multi-Dconv Head Transposed Attention (MDTA) and Gated-Dconv Feedforward Network (GDFN) modules perform the feature extraction task while reducing computational overhead. A masked auto-encoder(MAE) was used to process multiscale feature maps and to learn both local and global information. A combined loss function of Binary Cross-Entropy Loss (BCELoss) and Dice Loss deals with imbalance in classes as well as the overall strucmre of water bodies, particularly irregular or fragmented areas. The MHNet model was trained on several publicly available datasets such as the Tibetan Plateau Dataset (TPD), Gaofen Image Dataset (GID), WHDLD, DLRSD, and ISIC2018 dataset. On the TPD dataset, MHNet achieved the highest inIoU (94.45%), recall (96.47%), precision (97.72%), and Fl-score (97.03%).
(Chen et al., 2024) put forward a hybrid framework of CNN-Transformer, which is Lake Extraction Former (LEFormer). The architecture has mainly three modules: CNN Encoder (CE), Transformer Encoder (TE), and Cross-Encoder Fusion (CEF). The CE captures accurate fine-scale details and local spatial information. The TE acquires global context and long-range dependencies using a lightweight design with overlapped patch merging and a convolution layer for positional information. The CEF fuses fine-grained features from the CE with scene-level feamres from the TE via pointwise convolution, thereby improving mask prediction accuracy. The LEFormer was trained on two publicly available datasets, the Surface Watert SW) and Qinghai-Tibet Plateau Lake(QTPL) datasets.

4.4 Loss functions in Deep Learning Models

The cost function is critical for the optimization of DL models. Loss function calculates the error between the model's output and the reference label. In RSI, water areas occupy only a very small region of the scene, resulting in a class imbalance across many satellite images. If this issue is not handled properly, the model tends to prioritize land areas and miss smaller water regions. Choosing an appropriate loss function directly impacts how well the model accurately delineates water boundaries, particularly when dealing with class imbalance and noisy labels. The following section briefly explains these loss functions and their role.

1. **Binary Cross-Entropy(BCE) Loss:** Binary cross-entropy is the widely accepted cost function of the SWBE problem. It is highly suitable for pixel-level segmentation tasks and can be defined as:

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

where N is the batch size of images, y_i is the ground truth label, and $\hat{y}_i$ is the predicted label. This loss function works well when the dataset is balanced, but it can underperform when water pixels are sparse.

2. **Dice Loss:** Dice Loss is derived from the Dice Coefficient (DC). It measures the similarity between predicted and ground truth masks. It is especially useful when there is a significant class imbalance.

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_i y_i \hat{y}_i}{\sum_i y_i + \sum_i \hat{y}_i}$$

where y_i, denotes the ground truth, $\hat{y}_i$ represents the predicted value. This loss emphasizes overlap between predicted and actual water regions, making it suitable for detecting small or narrow water bodies.

3. **Jaccard Loss:** It is also called the Intersection over Union(IoU) Loss. This function penalizes non overlapping regions and helps in improving boundary precision.

$$\mathcal{L}_{JaccardLoss} = 1 - \frac{\sum y_i \hat{y}_i}{\sum y_i + \sum \hat{y}_i - \sum y_i \hat{y}_i}$$

where y_i denotes the ground truth, $\hat{y}_i$ represents the predicted value. IoU loss is particularly beneficial when the segmentation requires a high degree of spatial accuracy, which is critical in hydrological mapping.

4. **Focal Loss:** This loss function deals with the challenge of class imbalance. It mainly focuses on the hard, misclassified pixels and down weights the contribution of easy samples.

$$\mathcal{L}_{Focal} = -\alpha_t (1 - \hat{y}_t)^{\gamma} \log(\hat{y}_t)$$

It has two hyper parameters α_t manages the importance of positive and negative classes and γ mainly focuses on learning critical samples.

5. **Combination of different loss functions:** To leverage the strengths of different loss functions, researchers often combine BCE with Dice or IoU losses:

$$\mathcal{L}_{Combo} = \lambda \mathcal{L}_{BCE} + (1 - \lambda)\mathcal{L}_{Dice}$$

where k is a weight factor that maintains the share of Binary Cross Entropy loss and Dice loss. This hybrid technique aids models in learning both global pixel-wise accuracy and local region overlap, leading to more accurate outcomes.

The loss function must match the morphological characteristics of water bodies to be targeted. Among these, the Binary Cross-Entropy (BCE) loss function is the most common and widely used because of its ease of computation. But it fails in the cases where water pixels constitute a small percentage of the image. In the case of class-imbalanced scenarios, Dice Loss and Jaccard Loss are appropriate choices. The Tversky Loss extends overlap-based optimization with a distinction in FP (false positive) and FN (false negative) penalities, which is particularly useful for highly imbalanced data where preserving the information of the water pixels is crucial. Focal Loss tackles the other major problem in SWBE, where those hard-to-classify pixels are given more weight to increase the sensitivity of detecting narrow and fragmented water bodies. Recent studies have also shown that hybrid loss functions, like BCE and Dice Loss, offer a balanced optimization approach by jointly improving pixel-wise classification accuracy and region-wise segmentation accuracy.

5. Publicly Available Datasets

High-quality, annotated datasets of RSI are a basic requirement for algorithm development and validation. In recent years, many publicly available datasets have been presented, providing a broad range of spatial resolutions, image types (RGB, multispectral), and ground truth masks. This section describes frequently adopted datasets for SWBE.

(i) **WHDLD Dataset:** The Wuhan Dense Labeling Dataset (WHDLD) (Shao et at., 2020) is a recent and high-quality dataset designed specifically for tasks like LULC and semantic segmentation. It comprises high-resolution RGB images of Wuhan city acquired from numerous satellite sensors. The WHDLD was developed by researchers associated with Wuhan University in 2018. The dataset offers significant diversity in terms of geography, seasonal variation, and landscape complexity, making it suitable for deep learning-based segmentation models.

(ii) **Landsat Series(Landsat-8** and **Landsat-9):** The Landsat program, operated jointly by NASA and the USGS, provides freely accessible multispectral satellite imagery at a spatial resolution of 30 meters. Landsat datasets have been extensively used for water body detection through both spectral index-based approaches (e.g., NDWI, MNDWI) and ML models. Due to their long historical archive **(since** the 1970s), these datasets prove especially valuable for long-standing surface water monitoring and temporal panem assessment. Though they offer worldwide coverage, their comparatively coarse spatial resolution may constrain their utility in detecting small or narrow water bodies. The Global Surface Water Explorer (GSWE) is a *well*-known dataset derived from Landsat satellite imagery.

(iii) **Sentinel Dataset:** The Sentinel satellites launched under the European Union's Copernicus program provide powerful tools for water body analysis, especially through Sentinel-1 and Sentinel-2. The C-band SAR instrument on Sentinel-1 acquires images independent of atmospheric conditions and daylight, making it highly valuable for monitoring of surface water bodies in cloud-covered and monsoon-affected areas. On the other hand, the Multi-Spectral Scanner (MSS) board on Sentinel-2 gives image data in 13 spectral bands, including visible, red edge, and NIR bands, at spatial resolutions of 10, 20, and 60 meters. Furthermore, in recent years, some attempts have also been made to create hybrid datasets, which are a combination of SAR (Sentinel-1) and optical (Sentinel-2) sensors. (Wieland *et* al., 2023) launched a new global reference dataset, the SlS2-water dataset, which comprises 65 image triplets and DEM. It integrates radar and optical observations to facilitate floodwater and surface water detection.

(iv) **Deep Globe Water body Segmentation Dataset:** One significant dataset is the DeepGlobe Water Segmentation Dataset, which originates from the broader DeepGlobe 2018 Satellite Image Understanding (Demir et al., 2018). This dataset can also be used successfully for extracting water bodies. The images are obtained from high-resolution Digital Globe satellites whose spatial resolution is around 50 cm per pixel. Each image tile of this dataset has a size of 2448 × 2448 pixels, which is quite enough to provide sufficient details to delineate small and large water body features.

(v) **MODIS Dataset:** Moderate Resolution Imaging Spectroradiometer (MODIS) is a significant tool on Terra and Aqua satellites of NASA. It records images in 36 spectral bands of visible to thermal infrared, and its spatial resolutions are 250 in, 500 m, and 1 km, depending on the spectral band. Its revisit cycle has enabled it to capture almost daily global coverage, making it a very good resource in monitoring temporal variations in water bodies, especially at regional and continental levels. The MODIS data have found the widest applications in the production of global surface water products like the MODIS Water Mask (MOD44W).

These datasets vary depending on image type, spatial resolution, patch size, geographic coverage, and annotation method. Table 6 lists the most widely used datasets, including their characteristics and data modalities for SWBE.

5.1 Evaluation Metrics for DL Models

The quantitative assessment of DL models for binary water body segmentation is vital to determine how accurately a model can differentiate between water and non-water pixels. These parameters are derived from the confusion matrix, consisting of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). From these components, several indicators are measured to analyze the model's performance.

(i) **Overall Accuracy:** The OA is the proportion of correctly classified pixels—both water and non-water, out of the total number of pixels. Although it provides a general measure of performance, it may produce misleading results for imbalanced datasets dominated by non-water pixels.

$$OA = \frac{TP + TN}{TP + TN + FP + FN}$$

(ii) **Precision:** Precision represents the fraction of pixels predicted as water that are actually water. It helps assess the ability of the model to reduce false alarms, especially in cases where shadows, dark soil, or vegetation might be incorrectly classified as water.

$$Precision = \frac{TP}{TP + FP}$$

(iii) **Recall:** Recall is also called sensitivity. It quantifies the proportion of actual water pixels correctly detected by the model and measures the number of negative samples. A high recall indicates that most water regions are successfully identified.

$$Recall = \frac{TP}{TP + FN}$$

(iv) **Intersection over Union (IoU):** The Intersection over Union (IoU), also called the Jaccard Index, computes the ratio of overlapping area to total area between the predicted and actual regions of the water class. It evaluates how much the two areas intersect compared to their combined area. For binary segmentation, the Mean Intersection over Union (mIo provides an average measure of IoU across both water and non-water classes.

$$IoU = \frac{TP}{TP + FP + FN}$$

(v) **Mathews Correlation Coefficient:** The MCC is often called a balanced measure because it remains informative even when the classes are highly imbalanced-a common issue in water body segmentation, where the non-water class dominates the image.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

6. Challenges and Open Issues

There are certain challenges in detecting surface water bodies from RSI. These issues are due to data limitations, environmental factors, model generalization, and computational constraints. The first major issue is a lack of labeled datasets. Many DL models require large quantities of well-labeled data to produce precise segmentation. Manual labeling of water regions in RSI is time-consuming and requires domain expertise. Because of this, there are only a few large publicly available datasets that restrict the training and validation of ML and DL models. Another shortcoming is that high-resolution satellite imagery is not freely accessible. Although medium-resolution datasets, such as Landsat and Sentinel, are commonly adopted owing to their open accessibility, their spatial resolution (10-30 m) hinders the precise delineation of thin or discontinuous surface water bodies, particularly in complex topographies and urban environments. On the other hand, commercial high-resolution datasets (e.g., WorldView, GeoEye) offer finer spatial detail but are costly and not freely available. The variability in spectral characteristics and spatial patterns of water features is another significant challenge. The water's

Table 4: Summary of DL architectures for SWBE, including satellites, study area, results, and findings

DL Models	Dataset / Image Type	Study Area	Results	Findings	Authors
RRF—DeconvNet	Self-created Google Earth RGB images	Suzhou and Wuhan, China	OA: 96.5%	Large receptive fields and redundant parameters make segmentation slower and more difficult.	*(Miao et al., 2018)*
Inception CNN	Self-created JZBay dataset	Jiaozhou Bay, Qingdao, China	PA:94.8% UA:95.4%	Efficiency limits real-time coastline monitoring.	*(Liu et al., 2019)*
DenseNet	Gaofen-1 satellite images	Poyang Lake, Jiangxi, China	Precision: 96.1%	Faster convergence than ResNet, VGG, and SegNet.	*(Wang et at., 2020)*
W-Net	Self-created dataset, Landsat-8 OLI	River and non-river scenes	IoU: 94.35%, F1-score: 95.09%	Asymmetric convolution lowers parameter count.	*(Tambe, Talbar and Chavan, 2021)*
Faster R-CNN	Sentinel-2 and Landsat-8 OLI	Egypt	Acc: 98.7% (S-2), 96.1% (L-8)	Good boundary localization but weak on small water bodies.	*(Gharbia, 2023)*
FCN-based model	VHR optical images, Gaofen-2	South Beijing	FI: 92%	Captures spatial and contextual details effectively.	*(L. Li et at., 2019)*
SR—SegNet	Lake and river dataset	Namsto Lake and nearby river, China	71% fewer parameters	Dilated convolution improves fine-detail and small-river detection.	*(Weng et at., 2020)*
U-Net based model	Sentinel-2 RGB	Public dataset	OA: 90%, Dice: 86%	Attention gates and residual blocks enhance segmentation.	*(Kadhim and Premaratne, 2023)*
AER U-Net	Sentinel-2 RGB	Public dataset	IoU: 94%	Improves boundary extraction in shadows and mountains.	*(Jonnala et at., 2025)*
AFU-Net	ALOS PALSAR, Sentinel-1 SAR	Rivers in Americas, Asia, Africa	OA: 96.34%	AC loss sharpens and optimizes boundaries.	*(Han et at., 2024b)*
MSAttU-Net	Gaofen-2 imagery	Nanjing urban area	ASCR: 99.98%	Handles shadow interference effectively.	*(Xie et al., 2025)*
MSGF-Net	Gaofen-1 and Sentinel-2	Yellow River, Henan	OA: 98.60% (GF-1), 98.22% (S-2)	Swin Transformer reduces complexity and boosts multi-scale features.	*(Miao et al., 2024)*
U-Net + Vision Transformer	Self-created SAR dataset	Poyang Lake, Jiangxi	Acc: 98.45%, mIoU: 95.41%	Fewer parameters with improved efficiency.	*(Ye et al., 2023)*
WB-Former	WHDLD dataset	Wuhan, China	Precision: 98.26%, F1: 98.18%	MRF and Rolle's theorem-based loss improves segmentation.	*(Tian et al., 2025)*

appearance in a satellite image changes significantly because of seasonal changes, depth, turbidity, the presence of vegetation, or different sun angles. Moreover, water is usually confused with spectrally similar features like shadows, asphalt, or dark rooftops, resulting in false positives. The mixed pixel problem is another important challenge, particularly at the boundary areas between land and water. The problem of mixed pixels is primarily caused by low spatial resolution and a data-level problem, which compromises the accuracy of boundaries and mapping. Thresholding or index-based methods are sensitive to such ambiguity, while even advanced deep learning models may struggle to distinguish mixed pixels when boundaries are unclear, or reflectance patterns vary due to turbidity or vegetation. Cloud cover and atmospheric interference are major factors affecting the quality of optical RSI. The main reasons for spectral distortions in satellite images are atmospheric scattering, haze, shadows, and continuous clouds, which cause misclassification of water regions. Though cloud masking and atmospheric correction methods have been developed, they are still imperfect, particularly for tropical regions with frequent cloud cover.

Table 5: Comparison of traditional and modern methods for surface water body detection in remote sensing imagery

Method	Description	Advantages	Limitations / Drawbacks	References
Visual Interpretation	Manual delineation of water bodies by human experts using high-resolution imagery.	Reliable when done by experts resolves ambiguities by human judgment.	Subjective and inconsistent. Time-consuming and less accurate. Not scalable for large regions.	*(Paine and Kiser, 2012b) (Li and Sheng,2012)*
Single Band Thresholding	Applies a threshold on one spectral band (e.g., NIR SWIR) to distinguish water.	Simple and fast. Works for clear water in simple scenes.	Low accuracy. Thresholds vary by scene. Difficult to obtain optimum threshold value. Fails in shadowed, turbid, or urban regions.	*(McFEETERS, 1996)* *(Xu, 2006)*
Spectral Indices (e.g., NDWI, MNDWI, AWEI, NDMI)	Combines multiple spectral bands into indices that highlight water and suppress soil/vegetation.	More robust than single-band methods. Easy to apply at large scales.	Sensitive to shadows, seasonality, and turbidity. Threshold mning required. Urban areas remain problematic.	*(McFEETERS, 1996) (Xu, 2006) (Feyisa et al., 2014)*
Machine Learning-based (e.g., k-means, ISODATA, FCM, SVM, RF, CART)	Uses supervised or unsupervised machine learning with spectral or texture-based features.	Captures nonlinear relationships and provides higher accuracy than thresholding methods.	Requires prior knowledge. Performance depends on features. Less effective on complex boundaries.	*(F. Zhang et at., 2018b)*
Deep Learning-based (CNN/FCN/U-Net)	Learns spatial-spectral features using convolutional networks for pixel-wise segmentation.	High accuracy. robust to shadows, clouds, and turbidity.	Requires large annotated datasets. Computationally expensive. Risk of overfitting.	*(Kadhim and Preinaratne, 2023)*
Hybrid Transformer-based Deep Learning (e.g., CNN + Transformer, Swin-U-Net, SegFormer hybrids)	Combines CNN local features with Transformer global context modeling.	Capmres fine boundaries and long-range dependencies. Better generalization.	Very high computational cost. Requires large datasets. Complex architectures.	*(Zhang et at., 2023) (Tian et at., 2025)*

The multi-sensor fusion techniques of optical and SAR information have potential solutions, although they come with other challenges, namely geometric misalignment, radiometric normalization, and computational burden.A further open issue is the generalization and transferability of models. Most algorithms perform well on the datasets or geographic locations where they were trained but do not work when applied to new locations, sensors, and different time periods. The adaptability of models is limited by differences in spatial resolution, sensor characteristics, and environmental conditions. The computational load and resource requirements of DL models also restrict their deployment in real-time or field conditions. These models require high-end GPUs, large processing power, large memory, and extensive training time.Another important concern is model interpretability, which is particularly important for deep learning systems that are considered black boxes. Understanding how and why a model identifies certain regions as water or non-water is really important for decision-making. Lastly, the detection of narrow and urban water bodies remains a very critical task. These water bodies' features usually cover a few pixels and exhibit spectral confusion with non-water objects such as shadows, roads, or rooftops. Most algorithms perform well on large, homogeneous water surfaces while failing to detect small ponds, drainage channels, or water in built-up areas.

Table 6: Summary of publicly available datasets commonly used for water body extraction in remote sensing imagery

Dataset	Image Type	Spat. Resolution	No. Image	Image Patch	URL Link	Ref.
WHDLD Dataset	RGB image	2 m	4,940	256x256	hnps://sites.google.com/view/zhouwx/dataset#h.p hQS2jYeaFpV0	*(Shao et al., 2020)*
Landsat 8 Satellite imagery	RGB image	30 m	----------	7711×7531	hnp://glovis.usgs.gov	*(Acharya et al., 2016)*
Sentinel-2 Water Edges Dataset (SWED)	Sentinel-2 MSI	10 m	28,224	256x256	hnps://openmldata.ukho.gov.uk/	*(Seale et al., 2022)*
S1S2-Water Dataset	SAR image and MSS image	SAR 10-20m, Optical 10 m	65 triplets	256 x256	hnps://zenodo.org/records/11278238	*(Wieland et al., 2023)*
AIWR (Aerial Image Water Resources	Aerial RGB	50 m	800	650x650	hnps://data.mendeley.corn/datasets/d73rnpc529b/3	*(Noppitak, Gonwirat and Surinta, 2020)*
Water Bodies Dataset	RGB image	10 m	2,841	256x256	hnps://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies	-------
GLH-Water	VHR RGB image	0.3 m	250	12,800 x12,800 (cropped to 512x512)	hnps://jack-bo1220.github.io/project/GLH-water.html	*(Li et al., 2024)*
TPD (Tibetan Plateau)	Google Eanh RGB image	2 m	6,774	256x256	hnps://www.ncdc.ac.cn/portal/rnetadata'b4d9fb27-ec93-433d-893a-2689379a3fc0	(*MSLWENet: A Novel Deep Learning Network for Lake Water Body Extraction of Google Remote Sensing Images*)
GID	RGB image	0.8 m	150	6800×7200	https://captain-whu.github.io/GID/	(Tong *et al.*, 2020)
DLRSD	Aerial RGB image	0.3 m	2,100	256×256	https://sites.google.com/view/zhouwx/dataset#h.p_hQS2jYeaFpV0	(*Performance Evaluation of Single-Label and Multi-Label Remote Sensing Image Retrieval Using a Dense Labeling Dataset*)

7. Future Scope

In this article, we have emphasized four approaches for SWBE: visual interpretation, thresholding, machine learning, and deep learning. Each method is critically studied and analyzed. The following section elaborates on these future directions in detail. A recent trend is the combination of multimodal remote sensing information like optical, SAR, LiDAR, and thermal imagery. The multisource information can be used to mitigate the shortcomings of individual sensors, such as cloud cover in satellite images, and enhance modal fidelity under challenging conditions. Nonetheless, this demands better developments in model architecture capable of processing heterogeneous data types and aligning them spatially and temporally. Furthermore, the development of light and efficient models suitable for real-time or edge deployment is essential for operational monitoring. In the context of data, the development of standardized, high-resolution, and globally diverse annotated datasets for SWBE remains essential. Community-driven initiatives and open challenges could help accelerate benchmarking. A potential direction is the development of strong and generalizable models that can be utilized in diiTerent geographies, seasons, and sensor properties. The existing models are usually prone to overfitting problems to particular datasets or locations, limiting their applicability in unseen environments. The future effort should aim at the development of domain-adaptive models. Another key area is the explainability and interpretability of deep learning models. Most existing approaches work as "black boxes," rendering their decision-making mechanisms difficult to understand. Building models that can be explained using explainable AI (XAI) could improve user trust and adoption, especially in important applications like disaster response or policy formulation.

8. Conclusion

The extraction of water bodies is important for numerous application fields such as environmental monitoring, hydrological, and water resource management. Accurate and timely extraction of surface water bodies from RSI supports flood monitoring, irrigation planning, and sustainable water resource governance. In the last three decades, various traditional and automated approaches have been explored, including thresholding, spectral indices, machine learning algorithms, and deep learning frameworks and their hybrid architecmre. However, these techniques still face limitations in handling complex terrains, mixed pixels, seasonal variability, and generalization across diverse geographies. The latest developments in DL models have greatly changed the landscape of SWBE. CNNs, FCNs, transformer-based architectures, and hybrid or multi-branch models have significantly improved accuracy, robustness, and adaptability in segmenting water feamres from RSI. Despite these advancements, several challenges remain, such as computational costs, domain generalization, limited annotated datasets, and sensitivity to cloud cover or seasonal shifts. In addition, the black-box nature of deep learning creates interpretability challenges in critical decision-waking situations. Addressing these challenges through domain adaptation, lightweight architectures, explainable AI, and multi-modal fusion is an active research direction.This review article presents a structured overview of the evolution of SWBE methods. It also focuses on the key datasets, performance metrics, and loss functions. The prospect must focus on improving model interpretability, expanding annotated datasets, and integrating multi-source information to build more generalizable models.

Declarations

• Funding: No funds, grants, or other financial support were received during the preparation of this manuscript.
• Competing interests: There are no conflicts of interest related to work.
• Ethics approval: Not applicable.
• Consent for publication: Not applicable
• Consent to Participate: Not applicable
• Data availability: Not applicable

References

1. Acharya, T.D. et al. (2016) “Identification of water bodies in a Landsat 8 OLI image using a J48 decision tree,” Sensors, 16(7), p. 1075.
2. Adler-Golden, S. et al. (1998) “FLAASH, a MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations,” Proc. 7th Ann. JPL Airbome Earth Science Workshop. JPL Publication Pasadena, CA,pp.9-14.
3. Amitrano, D. et al. (2024) “Flood detection with SAR: A review of techniques and datasets,” Remote Sensing, 16(4), p. 656.
4. Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE transactions on pattern analysis and machine intelligence, 39(12), pp. 2481-2495.
5. Ball, G.H. and Hall, D.J. (1965) “ISODATA, a novel method of data analysis and pattern classification.” Available at: https://apps.dtic.mil/sti/html/tr/AD0699616/
6. Bangira, T. et al. (2019) “Comparing thresholding with machine learning classifiers for mapping cornplex water,” Remote Sensing, 11(11), p. 1351.
7. Belgiu, M. and Drăguţ, L. (2016) “Random forest in remote sensing: A review of applications and future directions,” *ISPRS journal of photogrammetry and remote sensing*, 114, pp. 24–31.
8. Berk, A. et al. (2009) “Recent developments in the MODTRAN% atmospheric model and implications for hyperspectral compensation,” 2009 IEEE International Geoscience and Remote Sensing Symposium. IEEE, p. II—262. Available at: https://ieeexplore.ieee.org/abstract/document/5418058/
9. Bezdek, J.C. (1980) “A convergence theorem for the fuzzy ISODATA clustering algorithms,” IEEE transactions on pattern analysis and machine intelligence, (1), pp. 1-8.
10. Bezdek, J.C. (2013) Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media.
11. Bijeesh, T.V. and Narasimhamurthy, K.N. (2020) “Surface water detection and delineation using remote sensing images: a review of methods and algorithins,” Sustainable Water Resources Management, 6(4), p. 68. Available at: https://doi.org/10.1007/s40899-020-00425-4.
12. Campbell, J.B. and Wynne, R.H. (2011) Introduction to remote sensing. Guilford press.
13. Chavez Jr, P.S. (1988) “An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data,” Remote sensing of environment, 24(3), pp. 459—479.
14. Chen, B. et al. (2024) “LEFormer: A hybrid CNN-transformer architecture for accurate lake extraction from rernote sensing imagery,” ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 5710-5714.
15. Chen, L.-C. et al. (2017) “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” IEEE transactions on pattern analysis and machine intelligence, 40(4), pp. 834-848.
16. Choubin, B. et al. (2017) “Watershed classification by remote sensing indices: A fuzzy c-rneans clustering approach,” Journal of Mountain Science, 14(10), pp. 2053-2063. Available at: hnps://doi.org/10.1007/s11629-017-4357-4.
17. Crist, E.P. and Cicone, R.C. (1984) “A physically-based transformation of Thematic Mapper data—The TM Tasseled Cap,” IEEE Transactions on Geoscience and Remote sensing, (3), pp. 256-263.
18. Dai, Z. et al. (2021) “Coatnet: Marrying convolution and attention for all data sizes,” Advances in neural information processing systems, 34, pp. 3965-3977.
19. Demir, I. et al. (2018) “Deepglobe 2018: A challenge to parse the earth through satellite images,” Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 172-181.
20. Dosovitskiy, A. et al. (2020) “An image is worth l6x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929 [Preprint]. Available at: https://arxiv.org/pdf/2010.11929/100.

21. Drusch, M. et al. (2012) “Sentinel-2: ESA's optical high-resolution mission for GMES operational services,” Remote sensing of Environment, 120, pp. 25-36.
22. European Space Agency (2026a) “Sentinel-1, RADARSAT-1, ENVISAT, TerraSAR-X Mission.” ESA Earth Online. Available at: hnps://earth.esa.int/eogateway/missions.
23. European Space Agency (2026b) “SPOT 6 Mission, WorldView-3 Mission.” ESA Earth Online. Available at: https://earth.esa.int.
24. Farooq, B. and Manocha, A. (2025) “Small water body extraction in remote sensing with enhanced CNN architecmre,” Applied Soft Computing, 169, p. 112544.
25. Feng, W. et al. (2018) “Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model,” IEEE Geoscience and Remote Sensing Leners, 16(4), pp. 618—622.
26. Feyisa, G.L. et al. (2014) “Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery,” Remote sensing of environment, 140, pp. 23-35.
27. Filatov, D. and Yar, G.N.A.H. (2022) “Forest and Water Bodies Segmentation Through Satellite Images Using U-Net.” arXiv. Available at: hnps://doi.org/10.48550/arXiv.2207.11222.
28. Frazier, P.S. and Page, K.J. (2000) “Water body detection and delineation with Landsat TM data.” Available at: https://www.cabidigitallibrary.org/dot/full/10.5555/20013008034 (Accessed: July 1, 2026).
29. Fu, J., Wang, J. and Li, J. (2008) “Study on the automatic extraction of water body from TM image using decision tree algorithm,” International Symposium on Photoelectronic Detection and Imaging 2007: Related Technologies and Applications. SPIE, pp. 17-25. Available at: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/6625/662502/Study-on-the-automatic-extraction-of-water-body-from-TM/10.1117/12.790602.
30 Gao, B.-C. (1996) “NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space,” Remote sensing of environment, 58(3), pp. 257-266.
31. Gao, H., Birken, C. and Lettenmaier, D.P. (2012) “Global monitoring of large reservoir storage from satellite remote sensing,” Water Resources Research, 48(9).
32 Gascon, F. et al. (2017) “Copernicus Sentinel-2A calibration and products validation status,” Remote Sensing, 9(6), p. 584.
33. Gasparovic, M. and Singh, S.K. (2023) “Urban surface water bodies mapping using the automatic k-means based approach and sentinel-2 imagery,” Geocarto International, 38(1), p. 2148757. Available at: https://doi.org/10.1080/10106049.2022.2148757.
34. Gautam, S. and Singhai, J. (2024) “Critical review on deep learning methodologies employed for water-body segmentation through remote sensing images,” Multimedia Tools and Applications, 83(1), pp. 1869-1889. Available at: https://doi.org/10.1007/s11042-023-15764-5.
35. Ge, C., Xie, W. and Meng, L. (2022) “Extracting lakes and reservoirs from GF-1 satellite imagery over China using improved U-Net,” IEEE Geoscience and Remote Sensing Letters, 19, pp. 1-5.
36. Gharbia, R. (2023) “Deep Learning for Automatic Extraction of Water Bodies Using Satellite Imagery,” Journal of the Indian Society of Remote Sensing, 51(7), pp. 1S11-1521. Available at: https://doi.org/10.1007/s12524-023-01705-0.
37. Gleick, P.H. (1993) Water in crisis. New York: Oxford University Press. Available at: http://www.ibiologia.unam.mx/pdf/directorio/z/introduccion/world watershed re.pdf.
38. Han, B. et al. (2024a) “AFUNet with active contour loss for water body detection in SAR imagery,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, pp. 17540-17554.
39 Han, B. et al. (2024b) “AFUNet with active contour loss for water body detection in SAR imagery,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, pp. 17540-17554.
40. Hirabayashi, Y. et at. (2013) “Global flood risk under climate change,” Nature climate change, 3(9), pp. 816-821.
41. Hong, S. et at. (2015) “Water area extraction using RADARSAT SAR imagery combined with landsat imagery and terrain information,” Sensors, 15(3), pp. 6652—d667.
42. Huang, C., Davis, L.S. and Townshend, J.R.G. (2002) “An assessment of support vector machines for land cover classification,” International Journal of Remote Sensing, 23(4), pp. 725-749. Available at: https://doi.org/10.1080/01431160110040323.
43. Huang, X. et al. (2015) “Combining pixel-and object-based machine learning for identification of water-body types from urban high-resolution remote-sensing imagery,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(5), pp. 2097-2110.
44. Ji, L., Zhang, L. and Wylie, B. (2009) “Analysis of dynamic thresholds for the normalized difference water index,” Photogrammetric engineering & remote sensing, 75(11), pp. 1307-1317.
45. Jonnala, N.S. et al. (2025) “AER U-Net: Anention-enhanced multi-scale residual U-Net structure for water body segmentation using Sentinel-2 satellite images,” Scientific Reports, 15(1), p. 16099.
46. Kadhim, I.J. and Premaratne, P. (2023) “A Novel Deep Learning Framework for Water Body Segmentation from Satellite Images,” Arabian Journal for Science and Engineering, 48(8), pp. 10429-10440. Available at: https://doi.org/10.1007/s13369-023-07680-5.
47. Kernker, R., Salvaggio, C. and Kanan, C. (2018) “Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning,” ISPRS journal of photogrammetry and remote sensing, 145, pp. 60-77.
48. Kothari, Siddharth et al. (2026) “Adversarial Robustness of Deep Learning Models for Inland Water Body Segmentation from SAR Images,” .

49. Lacaux, J.P. et at. (2007) “Classification of ponds from high-spatial resolution remote sensing: Application to Rift Valley Fever epidemics in Senegal,” Remote sensing of environment, 106(1), pp. 66-74.
50. Lee, H. et al. (2023) “IPCC, 2023: Climate Change 2023: Synthesis Report: A Report of the Intergovernmental Panel on Climate Change.” Available at: https://researchportalplus.anu.edu.au/en/publications/ipcc -2023-climate-change-2023-synthesis-report-a-report-of-the-in/ (Accessed: June 30, 2026).
51. Li, J. and Narayanan, R.M. (2003) “A shape-based approach to change detection of lakes using time series remote sensing images,” IEEE transactions on geoscience and remote sensing, 41(11), pp. 2466-2477.
52. Li, J. and Sheng, Y. (2012) “An automated scheme for glacial lake dynamics mapping using Landsat imagery and digital elevation models: a case study in the Himalayas,” International Journal of Remote Sensing, 33(16), pp. 5194—5213. Available at: https://doi.org/10.1080/01431161.2012.657370.
53. Li, L. et al. (2019) “Water body extraction from very high spatial resolution remote sensing data based on fully convolutional networks,” Remote Sensing, 11(10), p. 1162.
54. Li, S. et al. (2013) “Derivation of 30-m-resolution water maps from TERRA/MODIS and SRTM,” Remote Sensing of Environment, 134, pp. 417—430.
55. Li, Y. et al. (2024) “GLH-Water: A Large-Scale Dataset for Global Surface Water Detection in Large-Size Very-High-Resolution Satellite Imagery,” Proceedings of the AAAI Conference on Artificial Intelligence, 38(20), pp. 22213-22221. Available at: https://doi.org/10.1609/aaai.v38i20.30226.
56. Li, Z. et al. (2019) “Multiscale features supported DeepLabV3+ optimization scheme for accurate water semantic segmentation,” Ieee Access, 7, pp. 155787-155804.
57. Lian, L. and Jianfei, C. (2011) “Spatial-temporal change analysis of water area in Pearl River Delta based on remote sensing technology,” Procedia Environmental Sciences, 10, pp. 2170-2175.
58. Lillesand, T., Kiefer, R.W. and Chipman, J. (2015) Remote sensing and image interpretation. John Wiley & Sons.

https://books.google.com/books?hl=en&lr=&id=AFHDCAAAQBAJ&oi=fnd&pg=PR12&dq=remote+sensing+and+image+interpretation&ots=0Evke-iw7s&sig=PlY97BqR79KvDgqjDja8lpP5ezw (Accessed: May 30, 2026).
59. Liu, X.-Y. et al. (2019) “Coastline extraction method based on convolutional neural networks—A case study of Jiaozhou Bay in Qingdao, China,” IEEE Access, 7, pp. 180281—180291.
60. Liu, Z. et at. (2021) “Swin transformer: Hierarchical vision transformer using shifted windows,” Proceedings of the IEEE/CVF international conference on computer vision, pp. 10012-10022. Available at: https://openaccess.thecvf.com/content/ICCV2021/html/Liu Swin Transformer Hierarchical Vision Transform er Using Shified Windows ICCV 2021.
61. Long, J., Shelhamer, E. and Darrell, T. (2015) “Fully convolutional networks for semantic segmentation,” Proceedings of the IEEE conference on computer vision and panem recognition, pp. 3431—3440. Available at: http://openaccess.thecvf.com/content cvpr 2015/htmfLong Fully Convolutional Networks 2015 CVPR ap er.html (Accessed: June 30, 2026).
62. Louis, J. et al. (2016) “Sentinel-2 Sen2Cor: L2A processor for users,” Proceedings living planet symposium 2016. Spacebooks Online, pp. 1—8. Available at: https://elib.dir.de/107381/ (Accessed: June 30, 2026).
63. Malinowski, R. et al. (2015) “Detection and delineation of localized flooding from WorldView-2 multispectral data,” Remote sensing, 7(11), pp. 14853-14875.
64. Masek, J.G. et at. (2006) “A Landsat surface reflectance dataset for North America, 1990-2000,” IEEE Geoscience and Remote sensing letters, 3(1), pp. 68-72.
65. Mason, D.C. et al. (2009) “Flood detection in urban areas using TerraSAR-X,” IEEE transactions on geoscience and remote sensing, 48(2), pp. 882-894.
66. Maxwell, A.E., Warner, T.A. and Fang, F. (2018) “Implementation of machine-learning classification in remote sensing: an applied review,” International Journal of Remote Sensing, 39(9), pp. 2784-2817. Available at: https://doi.org/10.1080/01431161.2018.1433343.
67. 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), pp. 1425-1432. Available at: https://doi.org/10.1080/01431169608948714.
68. Miao, R. et at. (2024) “A method of water body extraction based on multiscale feature and global context information,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, pp. 12138-12152.
69 Miao, Z. et al. (2018) “Automatic water-body segmentation from high-resolution satellite images via deep networks,” IEEE geoscience and remote sensing letters, 15(4), pp. 602—d06.
70. Mishra, K. and Prasad, P.R.C. (2015) “Automatic Extraction of Water Bodies from Landsat Imagery Using Perceptron Model,” Journal of Computational Environmental Sciences, 2015, pp. 1-9. Available at: https://doi.org/10.1155/2015/903465.
71. Mobley, C.D. (1994) “Light and water: radiative transfer in natural waters,” (No Title) [Preprint]. Available at: https://cir.nii.ac.jp/crid/1971149384771940757/holdings (Accessed: June 30, 2026).
72. MSLWENet: A Novel Deep Learning Network for Lake Water Body Extraction of Google Remote Sensing Images (no date). Available at: hnps://www.mdpi.corn/2072-4292/12/24/4140 (Accessed: July 3, 2026).
73. Nagaraj, R. et al. (2023) “Surface water mapping and volume estimation of Lake Victoria using Machine Learning Algorithms,” 2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT). IEEE, pp. 1-6. Available at: https://ieeexplore.ieee.org/abstract/document/10170600/ (Accessed: July 1, 2026).

74. Nagaraj, R. and Kumar, L.S. (2021) “Performance analysis of machine learning techniques for water body extraction,” 2021 IEEE Bombay Section Signature Conference (IBSSC). IEEE, pp. 1-6. Available at: https://ieeexplore.ieee.org/abstract/document/9673372/ (Accessed: July 2, 2026).
75. Nagaraj, R. and Kumar, L.S. (2024) “Extraction of Surface Water Bodies using Optical Remote Sensing Images: A Review,” Earth Science Informatics, 17(2), pp. 893-956. Available at: https://doi.org/10.1007/s12145-023-01196-0.
76. Nath, R.K. and Deb, S.K. (2010) “Water-body area extraction from high resolution satellite images-an introduction, review, and comparison,” International Journal of Image Processing (IJIP), 3(6), pp. 265-384.
77. Nedeljkovic, I. (2004) “Image classification based on fuzzy logic,” The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 34(30), pp. 3-7.
78. Noppitak, S., Gonwirat, S. and Surinta, O. (2020) “Instance Segmentation of Water Body from Aerial Image using Mask Region-based Convolutional Neural Network.” Proceedings of the 2020 The 3rd International Conference on Information Science and System. ICISS 2020: 2020 The 3rd International Conference on Information Science and System, Cambridge United Kingdom: ACM, pp. 61—d6. Available at: https://doi.org/10.1145/3388176.3388184.
79 Ouma, Y.O. and Tateishi, R. (2006) “A water index for rapid mapping of shoreline changes of five East African Rift Valley lakes: an empirical analysis using Landsat TM and ETM+ data,” International Journal of Remote Sensing, 27(15), pp. 3153-3181. Available at: https://doi.org/10.1080/01431160500309934.
80. Paine, D.P. and Kiser, J.D. (2012a) Aerial photography and image interpretation. John Wiley & Sons. Paine, D.P. and Kiser, J.D. (2012b) Aerial photography and image interpretation. John Wiley & Sons.
81. Paisitkriangkrai, S. et al. (2016) “Semantic labeling of aerial and satellite imagery,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(7), pp. 2868-2881.
82. Pal, M. and Mather, P.M. (2005) “Support vector machines for classification in remote sensing,” International Journal of Remote Sensing, 26(5), pp. 1007-1011. Available at: https://doi.org/10.1080/01431160512331314083.
83. Paul, A., Tripathi, D. and Duna, D. (2018) “Application and comparison of advanced supervised classifiers in extraction of water bodies from remote sensing images,” Sustainable Water Resources Management, 4(4), pp. 905-919. Available at: https://doi.org/10.1007/s40899-017-0184-6.
84. Pekel, J.-F. et al. (2016) “High-resolution mapping of global surface water and its long-term changes,” Nature, 540(7633), pp. 418—422.
85. Performance Evaluation of Single-Label and Multi-Label Remote Sensing Image Retrieval Using a Dense Labeling Dataset (no date). Available at: https://www.mdpi.com/2072-4292/10/6/964 (Accessed: July 3, 2026).
86. Pierdicca, N. et al. (2008) “Integrating physical and topographic information into a fuzzy scheme to map flooded area by SAR,” Sensors, 8(7), pp. 4151—4164.
87. Purnam, K.K., Prasad, A.D. and Ganasala, P. (2024) “Water indices for surface water extraction using geospatial techniques: a brief review,” Sustainable Water Resources Management, 10(2), p. 70. Available at: https://doi.org/10.1007/s40899-024-01035-0.
88. Ramos, L.T. and Sappa, A.D. (2024) “Multispectral semantic segmentation for land cover classification: An overview,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, pp. 14295—14336.
89. Rokni, K. et al. (2015) “A new approach for surface water change detection: Integration of pixel level image fusion and image classification techniques,” International Journal of Applied Earth Observation and Geoinformation, 34, pp. 226—234.
90. Ronneberger, O., Fischer, P. and Brox, T. (2015) “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in N. Navab et al. (eds.) Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. Cham: Springer International Publishing (Lecmre Notes in Computer Science), pp. 234-241. Available at: https://doi.org/10.1007/978-3-319-24574-4 28.
91. Sarp, G. and Ozcelik, M. (2017) “Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey,” Journal of Taibah University for Science, 11(3), pp. 381-391. Available at: https://doi.org/10.1016/j.jtusci.2016.04.005.
92. Seale, C. et al. (2022) “Coastline detection in satellite imagery: A deep learning approach on new benchmark data,” Remote Sensing of Environment, 278, p. 113044. Available at: https://doi.org/10.1016/j.rse.2022.113044.
93. Sekertekin, A. (2021) “A Survey on Global Thresholding Methods for Mapping Open Water Body Using Sentinel-2 Satellite Imagery and Normalized Difference Water Index,” Archives of Computational Methods in Engineering, 28(3), pp. 1335-1347. Available at: https://doi.org/10.1007/s11831-020-09416-2.
94. Shao, Z. et at. (2020) “Multilabel remote sensing image retrieval based on fully convolutional network,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, pp. 318-328.
95. Shen, L. and Li, C. (2010) “Water body extraction from Landsat ETM+ imagery using adaboost algorithm,” 2010 18th International conference on geoinformatics. IEEE, pp. 1—4. Available at: https://ieeexplore.ieee.org/abstract/document/5567762/ (Accessed: July 2, 2026).

96. Sherrah, J. (2016) “Fully Convolutional Networks for Dense Seinantic Labelling of High-Resolution Aerial Imagery.” arXiv. Available at: https://doi.org/10.48550/arXiv.1606.02585.
97. Singh, A. (1989) “Review Article Digital change detection techniques using rernotely-sensed data,” International Journal of Remote Sensing, 10(6), pp. 989-1003. Available at: hnps://doi.org/10.1080/01431168908903939.
99 Singh, K.K. and Singh, A. (2017) “Identification of flooded area from satellite images using Hybrid Kohonen Fuzzy C-Means sigma classifier,” The Egyptian Journal of Remote Sensing and Space Sciences, 20(1), pp. 147—155.
100. Sivanpillai, R. and Miller, S.N. (2010) “Improvements in mapping water bodies using ASTER data,” Ecological Inforinatics, *S‹l),* pp. 73-78.
101. Smith, L.C. (1997) “Satellite remote sensing of river inundation area, stage, and discharge: a review,” Hydrological Processes, 11(10), pp. 1427-1439. Available at: hnps://doi.org/10.1002/(SICI)1099-1085(199708)11:10%3C1427::AID-HYP473%3E3.0.CO;2-S.
102. Tambe, R.G., Talbar, S.N. and Chavan, S.S. (2021) “Deep multi -feamre leasing architecture for water body segmentation from satellite images,” Journal of Visual Communication and Image Representation, 77, p. 103141.
103. Tian, Y. et al. (2025) “WB-Former: A Hybrid Model of CNN & Transformer for Water Body Extraction in Cornplex Scènes,” IEEE Transactions on Geoscience and Remote Sensing [Preprint]. Available at: https://ieeexplore.ieee.org/abstract/document/11127079/ (Accessed: July 2, 2026).
104. Tong, X.-Y. et al. (2020) “Land-cover classification with high-resolution remote sensing images using transferable deep models,” Remote Sensing of Environment, 237, p. 111322. Available at: https://doi.org/10.1016/j.rse.2019.111322.
105. Trochim, E.D. et al. (2016) “Remote sensing of water tracks,” Earth and Space Science, 3(3), pp. 106-122. Available at: https://doi.org/10.1002/2015EA000112.
106. U.S. Geological Survey (2025) “Landsat 9 OLI, Sentinel-2 MSI.” Available at: https://www.usgs.gov/landsat-missions/landsat-9.
107. Vanhellernont, Q. (2019) “Adaptation of the dark spectrum fining atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives,” Remote Sensing of Environment, 225, pp. 175-192.
108. Vaswani, A. et al. (2017) “Attention is all you need,” Advances in neural information processing systems, 30. Available at: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (Accessed: July 1, 2026).
109. Vermote, E. et al. (2016) “Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product,” Remote sensing of environment, 185, pp. 46-56.
110. Vermote, E.F. et al. (1997) “Second simulation of the satellite signal in the solar spectrum, 6S: An overview,” IEEE transactions on geoscience and remote sensing, 35(3), pp. 675—d86.
111. Vignesh, T. and Thyagharajan, K.K. (2017) “Water bodies identification from multispectral images using Gabor filter, FCM and canny edge detection rnethods,” 2017 international conference on information communication and embedded systems (ICICES). IEEE, pp. 1-5.
112. Vörösmarty, C.J. et al. (2010) “Global tbreats to huinan water security and river biodiversity,” nature, 467(7315), pp. 555-561.
113. Wang, G. et al. (2020) “Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks,” Remote sensing, 12(5), p. 795.
114. Wang, M. et al. (2024) “QTU-Net: quatemion transformer-based U-Net for water body extraction of RGB satellite image,” IEEE Transactions on Geoscience and Remote Sensing, 62, pp. 1-16.
115. Wang, S. et al. (2015) “A simple enhanced water index (EWT) for percent surface water estimation using Landsat data,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(1), pp. 90-97.
116. Wang, S. et al. (2025) “Mhnet: A masked hybrid network for robust water body segmentation from aerial images,” IEEE Transactions on Geoscience and Remote Sensing [Preprint]. Available at: https://ieeexplore.ieee.org/abstract/document/11037726/ (Accessed: July 1, 2026).
117. Weng, L. et al. (2020) “Water areas segmentation from remote sensing images using a separable residual segnet network,” ISPRS international journal of geo-information, 9(4), p. 256.
118. Wieland, M. et al. (2023) “SlS2-Water: A global dataset for semantic segmentation of water bodies from Sentinel-1 and Sentinel-2 satellite images,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17, pp. 1084—1099.
119. Wulder, M.A. et al. (2019) “Current status of Landsat program, science, and applications,” Remote sensing of environment, 225, pp. 127-147.
120. Xie, E. et al. (2021) “SegFormer: Simple and efficient design for sernantic segmentation with transformers,” Advances in neural information processing systems, 34, pp. 12077-12090.
121. Xie, Y. et al. (2025) “MSAttU-Net: A water body extraction network for Nanjing that overcomes interference in shaded areas,” IEEE Geoscience and Remote Sensing Letters [Preprint]. Available at: https://ieeexplore.ieee.org/abstract/document/10884824/ (Accessed: July 2, 2026).
122. Xu, H. (2006) “Modification of normalised difference water index (NDWI) to enhance open water features in rernotely sensed imagery,” International Journal of Remote Sensing, 27(14), pp. 3025-3033. Available at: https://doi.org/10.1080/01431160600589179.
123. Yang, F. et al. (2020) “Applied method for water-body segmentation based on inask R-CNN,” Journal of Applied Remote Sensing, 14(1), pp. 014502—014502.
124. Ye, F. et al. (2023) “Water body segmentation of SAR images based on SAR image reconstruction and an improved UNet,” IEEE Geoscience and Remote Sensing Letters, 21, pp. 1-5.

125. Yousefi, P. et at. (2018) “Water-body segmentation in satellite imagery applying modified kernel kmeans,” Malaysian Journal of Computer Science, 31(2), pp. 143-154.
126. Yu, L. et al. (2017) “Convolutional neural networks for water body extraction from Landsat imagery,” International Journal of Computational Intelligence and Applications, **16(01),** p. 1750001.
127. Yuan, K. et at. (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, pp. 7422-7434.
128. Zhang, F. et al. (2018a) “A simple automated dynamic threshold extraction method for the classification of large water bodies from landsat-8 OLI water index images,” International Journal of Remote Sensing, **39(11),** pp. 3429—3451. Available at: https://doi.org/10.1080/01431161.2018.1444292.
129. Zhang, F. et al. (2018b) “A simple automated dynamic threshold extraction method for the classification of large water bodies from landsat-8 OLI water index images,” International Journal of Remote Sensing, **39(11),** pp. 3429—3451. Available at: https://doi.org/10.1080/01431161.2018.1444292.
130. Zhang, J. and Foody, G.M. (1998) “A fuzzy classification of sub-urban land cover from remotely sensed imagery,” International Journal of Remote Sensing, 19(14), pp. 2721-2738. Available at: https://doi.org/10.1080/014311698214479.
131. Zhang, T. et al. (2023) “Water body extraction of the Weibe River basin based on MF-SegFormer applied to Landsat8 OLI data,” Remote sensing, 15(19), p. 4697.
132. Zhang, X., Li, J. and Hua, Z. (2022) “MRSE-Net: Multiscale residuals and SE-anention network for water body segmentation from satellite images,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, pp. 5049-5064.
133. Zhang, Y. et al. (2018) “Automatic and unsupervised water body extraction based on spectral-spatial features using GF-1 satellite imagery,” IEEE Geoscience and Remote Sensing Leners, 16(6), pp. 927-931.
134. Zhong, H.-F. et al. (2022) “NT-Net: A semantic segmentation network for extracting lake water bodies from optical remote sensing images based on transformer,” IEEE Transactions on Geoscience and Remote Sensing, 60, pp. 1-13.
135. Zhu, X.X. et al. (2017) “Deep learning in remote sensing: A comprehensive review and list of resources,” IEEE geoscience and remote sensing magazine, 5(4), pp. 8-36.
136. Zhu, Z. and Woodcock, C.E. (2012) “Object-based cloud and cloud shadow detection in Landsat imagery,” Remote sensing of environment, 118, pp. 83-94.