scientific reports

Check for updates

OPEN

AER U-Net: attention-enhanced multi-scale residual U-Net structure for water body segmentation using Sentinel-2 satellite images

Naga Surekha Jonnala[1], Shaik Siraaj[2], Y. Prastuti[2], P. Chinnababu[2], B. Praveen babu[2], Shonak Bansal[3]✉, Prashant Upadhyaya[3], Krishna Prakash[1]✉, Mohammad Rashed Iqbal Faruque[4]✉ & K. S. Al-mugren[5]

The automatic segmentation of water bodies from remote-sensing satellite images offers valuable insights into water resource management, flood monitoring, environmental changes, and urban development. However, extracting water bodies from satellite imagery can be challenging due to factors such as varying water body shapes, diverse environmental conditions, cloud cover, and shadows. These difficulties have a significant impact on waterbody segmentation, particularly in precisely maintaining high-quality segmented images and determining the boundaries of waterbodies. To overcome these issues, researchers have introduced several approaches; however, they suffer from precisely identifying the boundaries of waterbodies due to their irregular shapes. This difficulty is particularly pronounced in traditional threshold-based and machine-learning techniques, which often struggle to achieve accurate segmentation when confronted with complex structures, cluttered backgrounds, or objects of varying sizes and shapes. The objective of this research is to develop innovative deep-learning (DL) approaches to address these challenges and enhance the accuracy of waterbody segmentation in Remote sensing applications. This research introduces a deep learning model, namely AER U-Net architecture, which integrates advanced architectural elements into U-Net, such as residual blocks, self-attention mechanisms, and dropout layer, due to which the model significantly enhances segmentation accuracy and generalization capability. The architecture employs a contracting path consisting of convolutional layers, batch normalization, and activation layers to extract multi-scale features. Residual blocks improve feature learning efficiency while addressing the vanishing gradient issue through the inclusion of skip connections. Dropout layers in the encoder and bottleneck paths are incorporated for regularization, reducing the risk of overfitting. Additionally, the attention mechanism ensures precise refinement of skip connections, further improving segmentation performance. The model is trained using the Adam optimizer combined with a binary cross-entropy loss function, making it highly effective for binary segmentation tasks with an IoU score of 0.94, highlighting its effectiveness for practical environmental applications.

Keywords Enhanced attention mechanisms, Deep learning, Residual blocks, Sentinel-2 satellite imagery, And water body segmentation

Water bodies, including rivers, lakes, reservoirs, and wetlands, play a critical role in sustaining ecosystems, supporting biodiversity, and meeting the water needs of human societies. Accurate mapping and monitoring of these water bodies are essential for effective water resource management, climate change adaptation, and disaster preparedness[1]. Satellite imagery, with its ability to cover large spatial extents and provide repeatable observations, has emerged as a powerful tool for identifying and analyzing surface water features. Advances in

[1]Department of Electronics and Communication Engineering, NRI Institute of Technology, Agripalli, Eluru, AP 521212, India. [2]Department of AI & ML, NRI Institute of Technology, Agripalli, Eluru, AP 521212, India. [3]Department of Electronics and Communication Engineering, Chandigarh University, Gharuan, Punjab, India. [4]Space Science Centre (ANGKASA), Institute of Climate Change (IPI), Universiti Kebangsaan Malaysia (UKM), 43600 Bangi, Selangor D. E., Malaysia. [5]Physics Department, Science College, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. ✉email: shonakk@gmail.com; k_krishna2k7@yahoo.co.in; rashed@ukm.edu.my

remote sensing technology and data availability, such as high-resolution imagery from Sentinel-2 satellites, have further enhanced the precision and efficiency of water body segmentation.

Sentinel-2, part of the Copernicus program initiated by the European Space Agency (ESA), offers multi-spectral imagery with a spatial resolution ranging from 10 to 60 m. Its spectral bands, spanning the visible, near-infrared (NIR), and shortwave infrared (SWIR) regions, provide valuable data for distinguishing water bodies from other land covers[2]. These bands make it possible to analyze surface reflectance properties and accurately delineate water features. However, the complexity of real-world conditions, such as the presence of vegetation, shadows, and atmospheric disturbances, poses significant challenges for accurate water body extraction from satellite imagery.

Conventional techniques for water body segmentation often rely on spectral indices. Among these, the Normalized Difference Water Index (NDWI)[3] and its modified versions, such as the Modified NDWI (MNDWI)[4] and Automated Water Extraction Index (AWEI), have been widely used[5]. These indices exploit the differences in reflectance between water and other surfaces in specific spectral bands, providing a straightforward method for water detection. While effective under controlled conditions, these methods are sensitive to external factors such as cloud cover, turbidity, and mixed pixels, leading to inaccuracies in complex or heterogeneous environments[6].

To overcome these limitations, researchers have increasingly turned to machine learning approaches, which leverage data-driven algorithms to enhance classification and segmentation accuracy[7]. Techniques such as Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting classifiers have been successfully applied to combine spectral, spatial, and textural features for water body segmentation. These models, which do not depend solely on pre-defined indices, demonstrate greater adaptability to varying environmental conditions and datasets. However, traditional machine learning approaches often require manual feature engineering, which can be time-consuming and limits scalability[8].

In recent years, advancements in artificial intelligence (AI) and deep learning (DL) have introduced a paradigm shift in remote sensing applications. Deep learning models, such as convolutional neural networks (CNNs), have shown remarkable performance in image analysis tasks, including semantic segmentation. These models learn hierarchical features directly from raw data, eliminating the need for manual feature design and enabling high precision across diverse scenarios. Furthermore, the flexibility of deep learning architectures allows them to adapt to complex patterns, such as the dynamic nature of water boundaries, seasonal changes, and environmental disturbances.

This study highlights the significance of utilizing advanced remote sensing techniques and deep learning models to tackle the challenges of water body segmentation. By incorporating multi-spectral satellite imagery and cutting-edge algorithms, the aim is to improve the accuracy, efficiency, and scalability of water body monitoring systems, ultimately contributing to a deeper understanding of global water dynamics and better resource management. In line with this objective, recent research has increasingly focused on deep learning-based models. In this context, we have reviewed several methodologies for extracting water body areas from satellite images.

Dmytro and Ghulam[9] proposed a U-Net-based model for segmenting water bodies from satellite images, achieving an Intersection over Union (IoU) score of 0.60. Tin Moh and Zin Mar [10] implemented a block attention-based U-Net approach to distinguish water and non-water regions in remote satellite images, achieving an IoU of 0.61. Silpalatha and Jayadeva[11] introduced a ResNet-based method for segmenting water bodies in satellite images with an IoU of 0.75.

Harika et al.[12] applied a DeepLabV3 + model to extract water-contaminated areas from color-based satellite imagery, achieving an IoU of 0.72. The semantic segmentation network (SegNet) model, used by Badrinarayanan et al.[13], was also employed to segment water bodies in satellite images, resulting in an IoU of 0.77. Finally, Paszke et al.[14] utilized the efficient neural network (ENet) model to segment water bodies, yielding an IoU of 0.79. "Table 1 represents different state-of-the-art Models."

From the above-mentioned literature, we address the following issues:

1. Despite the use of advanced models like U-Net, ResNet, and DeepLabV3+, the Intersection over Union (IoU) scores achieved by these models remain relatively modest, with values ranging from 0.60 to 0.79. This suggests that while the models are effective to some degree, they still struggle to achieve high levels of precision in complex satellite imagery.
2. Several models, including the DeepLabV3 + and ResNet-based approaches, may be sensitive to the quality and resolution of the satellite images. Variations in lighting conditions, weather patterns, or cloud cover can degrade the model's ability to accurately segment water bodies.

These challenges underscore the necessity for continued refinement and adaptation of deep learning models to achieve more reliable and efficient water body segmentation. To address these issues, we incorporated attention mechanisms and ResNet architectures into U-Net models. This integration enhances the model's ability to focus on key features, improves its generalization across diverse datasets, and boosts overall performance in water body segmentation tasks. AER U-Net, which stands for Attention-Enhanced Multi-Scale Residual U-Net, is a fully convolutional model used for semantic segmentation. Since AER U-Net offers the best overall performance with the least amount of information loss, it stands out among the many attempts to use current neural networks for image segmentation. To improve the fundamental prediction results, various scholars have improved the semantic segmentation results, especially in terms of better edge and boundary recognition. To get high-resolution predictions, we employed long-distance residual connections for multi-scale features throughout the downsampling procedure. Post-processing the segmentation findings is done by the Residual Refines Module, an independent encoder-decoder. We employ a Refinement Residual Block to improve feature maps. Global and local refinement are used in this special-purpose refinery network to increase forecast accuracy[23]. However, a lot

Authors [citations]	Year	Methodology	Limitations
Akiyama et al.,[15]	2021	Deep Learning with CNN SegNet architecture	This suggests that there is a need for further advancement and improvement in the abilities of a system or technique to generalize and understand a wide range of images. The current capabilities are not sufficient to handle various types of images effectively. Therefore, additional efforts, research, and development are necessary to enhance the system's capacity to process and interpret diverse images accurately. This could involve refining algorithms, expanding training data, and incorporating more sophisticated techniques to ensure that the system can generalize its understanding across a broader spectrum of visual content.
Moradkhani et al.,[16]	2022	deep stacked ensemble model	This method requires further improvement which fails in cluttered images.
Yuan et al.,[17]	2021	MC-WBDN	The analyses that have been conducted are not relevant or suitable for being applied to the current strategy being discussed. In other words, the findings, conclusions, or methods derived from other hydrological investigations cannot be effectively used in conjunction with the strategy under consideration. This could be due to differences in the context, objectives, parameters, or underlying assumptions of the strategy, rendering the findings from those other investigations unsuitable or unfeasible for direct application in this specific case.
Feng et al.,[18]	2018	Deep U-Net	This method requires further improvement which fails in irregular shapes.
Zhang et al.,[19]	2022	MRSE-Net	This method requires further improvement which fails in boundary detections.
Li et al.,[20]	2022	the pixel-based convolutional neural network method	In the pursuit of achieving accurate segmentation, the method described falls short in its ability to detect a multi-scale, visible light band within satellite images. This means that when applying this approach to segmenting images captured by satellites, it lacks the capability to effectively recognize and utilize a range of visible light wavelengths at different scales. As a result, the method might struggle to accurately delineate distinct areas or objects within the images based on their visible light characteristics, which can impact the precision and reliability of the segmentation process.
Chen et al.,[21]	2021	Feature pyramid enhancement and pixel pair matching	That the method being discussed doesn't contribute positively to the dataset's suitability for segmentation, implying that it doesn't address the dataset's shortcomings or enhance its quality to yield better segmentation outcomes.
Miao et al.,[22]	2018	RRF DeconvNet	This technique is applied to the rough annotations made by individuals, it does not result in any noticeable improvement or enhancement in the quality, accuracy, or clarity of those annotations. This lack of improvement might be due to limitations in the technique itself, the complexity of the annotations, or the specific characteristics of the rough annotations that make them resistant to enhancement through this method.

Table 1. Different state of Art models.

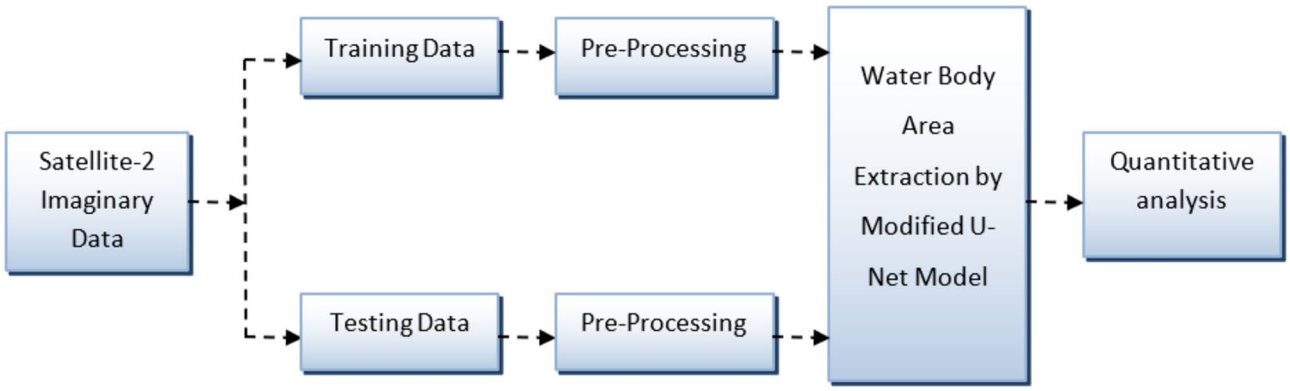


Fig. 1. Working process of the implemented approach.

of these methods have trouble reliably detecting water across wide areas, which frequently leads to unreliable or insufficient identification of water bodies in large-scale assessments. Our work's primary contribution is.

- AER U-Net, a multi-layered residual model intended for segmentation, is the suggested framework. Through the use of multi-scale residual networks, multi-dilated convolutions, and skip connections, AER U-Net integrates techniques to improve accuracy, allowing the model to capture complex data and increase segmentation precision.
- Utilising effective training methods and network architecture optimisation strategies, the model was trained with the Adam optimiser to increase performance. In order to accelerate convergence, the model also incorporates transfer learning. Data augmentation during training was one of the strategies used to expose the model to a variety of scenarios and improve its robustness. Additionally, regularisation strategies, careful hyperparameter selection, and exhaustive validation on a representative dataset all increased robustness.

Materials and models

Figure 1 presents a streamlined workflow for the segmentation of water bodies from Sentinel-2 satellite imagery, illustrating the sequence of key processes. It begins with the acquisition of Sentinel-2 imagery, capturing high-resolution multi-spectral data over the target regions. This raw data is then divided into training and testing datasets to facilitate the development and evaluation of the segmentation model. In the next stage, the data

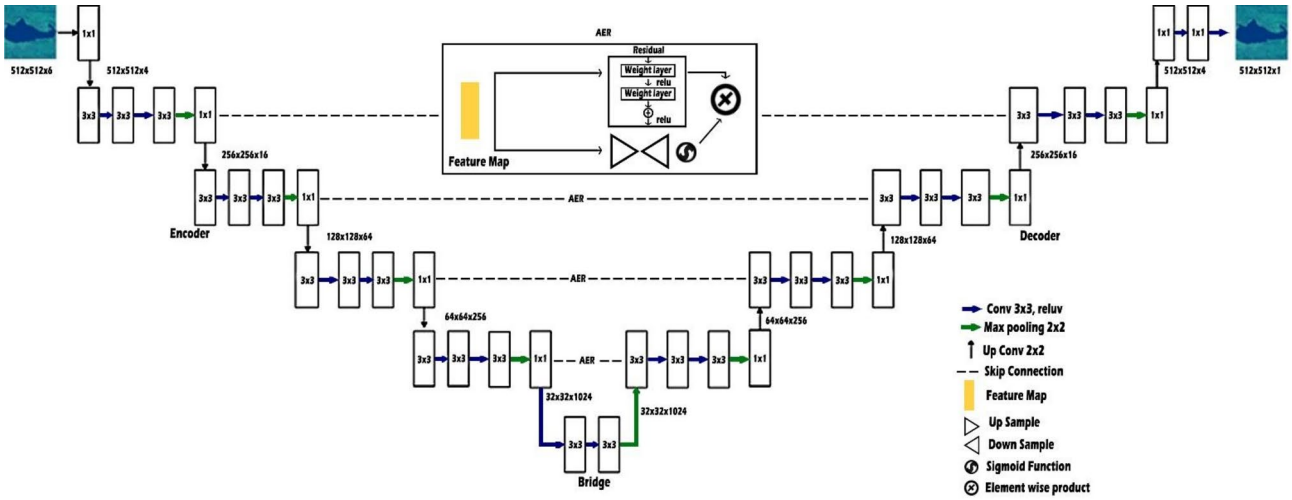


Fig. 2. Workflow of AER U-Nets.

(a) (b)

Fig. 3. Sample water body images: (**a**) Original; (**b**) Mask.

undergoes pre-processing to enhance its quality and ensure consistency. This step typically includes operations such as normalization and resizing to standard dimensions, which prepare the imagery for subsequent analysis.

Following pre-processing, the refined data is input into a modified U-Net model for segmentation. This advanced deep learning architecture is tailored to accurately identify and delineate water bodies within the imagery, leveraging both spatial and spectral information to deliver precise results. Finally, the segmented output is subjected to quantitative analysis, where performance metrics such as accuracy, precision, recall, and IIoU are computed. These metrics provide a rigorous evaluation of the model's performance, ensuring reliability and applicability in real-world scenarios. This comprehensive workflow highlights the integration of remote sensing, data preparation, advanced modeling, and evaluation to achieve accurate water body segmentation from satellite imagery.

Materials

The dataset utilized for model training and evaluation focuses on identifying and segmenting water bodies from satellite imagery. It is obtained from the Kaggle dataset titled Satellite Images of Water Bodies, featuring satellite data captured by the Sentinel-2 satellite[24]. The dataset is structured into two primary folders: Images and Masks. Here, masks are generated using the NWDI, a standard method employed to detect and map water bodies in satellite imagery. The NWDI exploits the spectral differences between water and non-water surfaces by comparing the reflectance in the green and NIR spectral bands. Water typically absorbs infrared wavelengths while reflecting green light, making it easier to detect using this index. The mathematical formula used for computing the NWDI is as follows:

$$\mathrm{NDWI} = \mathrm{Band3} - \mathrm{Band8}\ \mathrm{Band3} + \mathrm{Band8} \tag{1}$$

where, Band 3 represents Sentinel-2's green channel and Band 8 illustrates Sentinel-2's Near-Infrared (NIR) channels. "Figure 2 represents the workflow of AER U-Net", and "Figure 3 represents the sample water body images with the corresponding masks."

Pre-processing

Data preprocessing is an essential first stage in preparing data for DL model training. It involves cleaning, organizing, and converting raw data into a suitable format that allows seamless interaction with the model. The preprocessing phase focuses on discarding irrelevant information and addressing inconsistencies in the dataset.

Furthermore, it standardizes the data to ensure uniformity across all inputs, facilitating efficient and effective model training. The following sequence of operations is carried out during the preprocessing phase:

1. **Image Resizing**: The first stage involves resizing all images in the dataset to a uniform size. Working with smaller, uniformly sized images speeds up the training process and reduces computational time compared to processing larger and varied image sizes. The resizing is accomplished using the *resize* function from the cv2 module, ensuring each image adheres to the same dimensions for consistency.
2. **Pixel Scaling**: After resizing, each image is passed through a function called mask_split_threshold, which scales the pixel values of the images to a range between 0 and 1. This normalization ensures that the input data has a standard range, making it easier for the model to learn patterns without being influenced by varying raw value ranges. Scaling accelerates the convergence of models during the training process and improves overall performance.
3. **Image Padding**: The final step involves removing unnecessary padding bits that may exist around the images. These bits can lead to inconsistencies during both training and testing. All images must be identical in size for the model to perform optimally. Hence, this step ensures that only the relevant portions of the images are used, discarding any extraneous or unwanted pixels that may introduce noise or distortions.

AER U-Net

U-Net[25] is a well-known deep-learning approach for image segmentation due to its precision, robustness, and adaptability. Therefore, based on this idea, we proposed a modified U-Net by incorporating advanced building blocks like residual connections and attention mechanisms, making it more robust and capable of focusing on critical regions in satellite images. The basic building blocks of the suggested U-Net are described below, and its specifications are tabulated in Table 2.

1. **Convolution block**: The convolutional block is the core unit for feature extraction. It applies a convolutional layer followed by batch normalization and a non-linear activation function. Batch normalization ensures stability and accelerates training, while the activation function introduces non-linearity to model complex mappings. This block is strategically used throughout the network for feature extraction and transformation[26].
2. **Residual Block**: Residual blocks address the vanishing gradient problem by introducing skip connections. They preserve essential features across layers and allow deeper networks to train effectively. Each residual block aligns the input channels to match the output using a 1×1 convolution, followed by two convolutional blocks interleaved with a dropout layer for regularization. Finally, the shortcut connection adds the input back to the output, ensuring critical information is retained. This design facilitates better gradient flow and ensures that critical information is not lost as the network depth increases.
3. **Attention Block**: The attention block refines skip connections by focusing on relevant spatial regions. It aligns the skip connection and gating signal dimensions using 1×1 convolution. The feature maps are combined and activated with *ReLU*, followed by a *sigmoid* function to generate an attention map. This attention map highlights important regions, which are multiplied with the skip connection to refine the input for the decoder. This mechanism is particularly useful in tasks like water body segmentation, where distinguishing between the foreground (water) and background is crucial.
4. **Encoder**: The encoder captures hierarchical feature representations from the input image. Each level consists of a residual block for feature extraction, followed by a max-pooling layer to reduce spatial dimensions. The number of filters doubles with each level, enabling the network to learn increasingly abstract patterns. The encoder's role is to transform the input into a compressed, high-dimensional representation that retains essential features for segmentation.
5. **Bottleneck (Center)**: The bottleneck serves as the transition between the encoder and the decoder. It is designed to process the compressed features obtained from the encoder, extracting the deepest and most abstract representations of the input. This module consists of a residual block with 256 filters and an additional dropout layer for regularization. By capturing high-level contextual information, the bottleneck ensures that the decoder has access to features representing both global and local patterns in the input image.
6. **Decoder**: The decoder reconstructs the segmentation mask by upsampling feature maps and merging them with attention-refined skip connections. Each level begins with a transposed convolution to upsample the feature maps, followed by an attention block that combines the upsampled features with those from the encoder. A residual block processes the combined features to enhance detail and accuracy. This process is

Component	Filters	Kernel Size	Other Specifications
Convolutional Block	Configurable	3	Batch normalization, ReLU, and L2 regularization with weight decay of 0.0001
Residual Block	Configurable	3	Skip connections and Dropout = 0.3
Attention Block	Configurable	1	Activation: ReLU, and Sigmoid; Multiplicative attention.
Encoder Levels	32, 64, 128	3	MaxPooling2D, and Spatial downsampling
Bottleneck Filters	256	3	Dropout = 0.3, and Deep feature extraction
Decoder Levels	128, 64, 32	2	Transposed convolutions and Attention-enhanced skip connections
Output Layer	1	1	Activation: Sigmoid

Table 2. Block-wise configurations of the proposed modified U-Net.

repeated at each level, reducing the number of filters and progressively restoring the image's original resolution[26].

7. **Output**: The output layer produces the final segmentation mask, representing the probability of each pixel belonging to the target class. It uses a 1×1 convolution to reduce the feature maps to a single channel, followed by an activation function to normalize the predictions. The *sigmoid* activation ensures that the output values are in the range [0, 1], suitable for binary segmentation tasks. This design enables the network to produce precise segmentation masks with well-defined boundaries.

- Attention Enhanced U-Net Framework: The Attention Enhanced U-Net Framework is an enhanced U-Net architecture that incorporates attention methods to improve segmentation performance and fine-tune feature selection. Standard U-Net transfers encoder features straight to the decoder via skip connections, which could add extraneous data. By using self-attention mechanisms or attention gates (AGs), the network suppresses background noise and selectively focuses on the most pertinent areas, increasing segmentation accuracy. Transformer-based U-Net (TransUNet) and Attention U-Net with SE blocks are two variants that improve spatial and channel-wise feature learning even more. When accurate segmentation is essential, this method works especially well in autonomous systems, medical imaging, and remote sensing. In complicated segmentation tasks, the attention-enhanced model performs better because it decreases false positives, increases robustness, and guarantees more effective feature utilisation.
- Residual Blocks for Enhanced Feature Extraction: Residual Blocks for Enhanced Feature Extraction help deep neural networks learn complex features more effectively by addressing disappearing gradients. By avoiding one or more layers, a residual block is made up of shortcut connections, also known as skip connections, which enable the network to learn residual mappings rather than direct transformations. In addition to enabling deeper designs without deterioration, this helps maintain significant features. Residual blocks stabilise training and improve gradient flow by combining batch normalisation and ReLU activation. Due to their extensive use in architectures such as ResNet, U-Net variations, and Transformer models, they are especially good at deep segmentation, image processing, and medical imaging tasks because they enhance feature representation and capture minute details.
- Adaptive Adam Learning Optimization: The proposed approach leverages the advantages of important optimisation techniques to increase training efficiency and convergence by training the model using the Adaptive Adam optimiser with a learning rate of 0.001. RMSprop, momentum, and stochastic gradient descent (SGD) components are all incorporated into the Adaptive Adam optimiser, which expands upon the standard Adam optimiser. By modifying the learning rate for every parameter according to recent gradients, RMSprop helps to stabilise the optimisation and manage noisy gradients or shifting goals. Incorporating the momentum term speeds up convergence by minimising oscillations, smoothing updates across iterations, and facilitating the optimizer's rapid passage over shallow areas of the loss surface. The stochastic aspect of SGD, on the other hand, enables the model to update parameters using mini-batches, escaping local minima and perhaps producing more generalised solutions. Particularly in challenging deep learning problems, these elements work together to provide a more flexible and effective optimiser that converges more quickly and steadily. By maintaining a learning rate of 0.001, the approach avoids overfitting and guarantees significant advancement towards the ideal solution while striking a compromise between quick convergence and stability.

Quantitative measures

Quantitative measures for image segmentation provide metrics that allow you to objectively evaluate the quality of a segmentation model. These measures are essential for assessing how well the model partitions an image into meaningful regions. In this work, we considered Intersection over Union (IoU), dice, recall, precision, F1-score, and accuracy[28]. The entire process of the proposed model is illustrated in Algorithm 1.

Evaluation metrics

- **Accuracy**: It is defined as the fraction of the total count of appropriately categorized instances from the total count of instances.

$$\mathrm{Accuracy} = \mathrm{Tp} + \mathrm{TnTp} + \mathrm{Tn} + \mathrm{Fp} + \mathrm{Fn}$$

Tp -True Positive; **Tn** -True Negative; **Fp** -False Positive; **Fn** -False Negative.

- **Precision**: The ratio of appropriately categorized positive instances to the total count of positively predicted instances.

$$\mathrm{Precision} = \mathrm{TpTp} + \mathrm{Fp}$$

- **Recall**: The fraction of appropriately categorized positive instances from the total count of positive instances.

$$\mathrm{Recall} = \mathrm{TpTp} + \mathrm{Fn}$$

- **F1_Score**: It can be defined as the average harmonic between recall and precision.

$$\mathrm{F1Score} = 2\mathrm{RecallprecisionRecall} + \mathrm{precision}$$

Input: Satellite-2 based water body images

Output: Water body area extraction

Procedure

Step 1: Data collection

I = Water body images

M = Water body masks

Step 2: Data Pre-processing

Step 2a: Image resizing using *cv2.resize()* method

Step 2b: Image scaling based on *split_threshold* approach

Step 2d: Removal of padding values

Step 3: Construction of a Modified U-Net using specifications mentioned in Table 2

Step 3a: Construct a convolutional block

Step 3b: Construct a residual block

Step 3c: Construct an attention block

Step 3d: Construct an encoder block

Step 3e: Construct a bottleneck block

Step 3f: Construct a decoder block

Step 4: Train and compile the model using the Adam optimizer with a learning rate of 0.0001. Set the training to run for 50 epochs, employing the binary cross-entropy loss function along with *EarlyStopping* and *ReduceLROnPlateau* criteria for better optimization.

Step 5: Generate the predicted water body area images as the model's output.

Step 6: Conduct a quantitative analysis using the metrics outlined in the Quality Measures section.

Algorithm 1. Water Body Segmentation

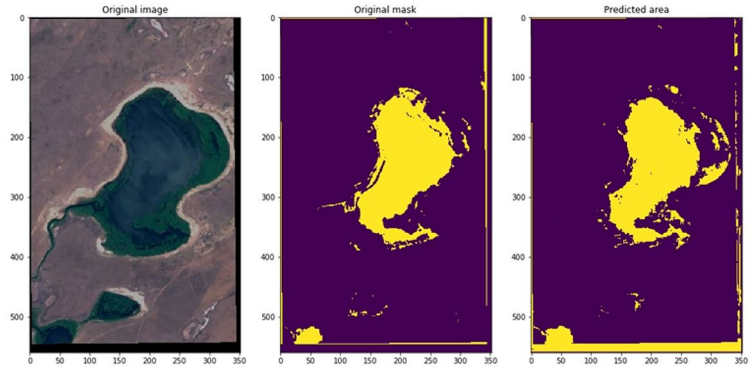
Results and discussion

"Table 3 represents Data set Description" and this section outlines the experimental findings of the proposed model, emphasizing its performance relative to leading methods in the field. A thorough examination of the results is provided, illustrating why our approach delivers superior outcomes compared to current techniques.

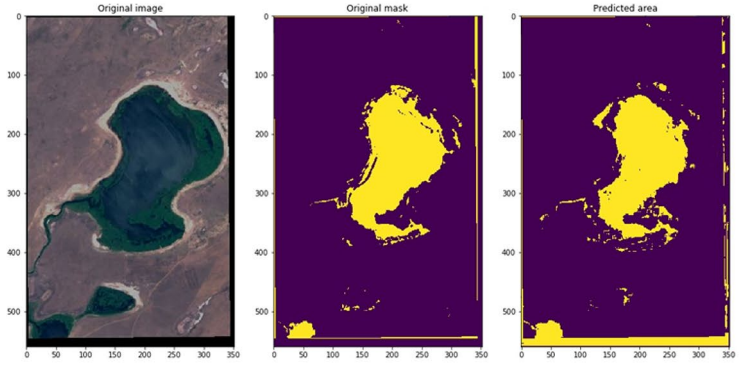
To predict water body areas from satellite images, we began by pre-processing the images through resizing, scaling, and padding. Next, we applied a modified U-Net model to detect water regions in the processed images[29]. The model learned relevant features automatically through multiple hidden layers and was trained using the backpropagation algorithm. Further model's performance was evaluated using various metrics, including IoU, Dice, precision, recall, F1-score, and accuracy, which are presented in Table 7. For the experiments, the dataset was divided into 80% for training and 20% for testing. The experiments were conducted on a desktop featuring

Data	Number of Images
Training	2272
Testing	568
Total	2840

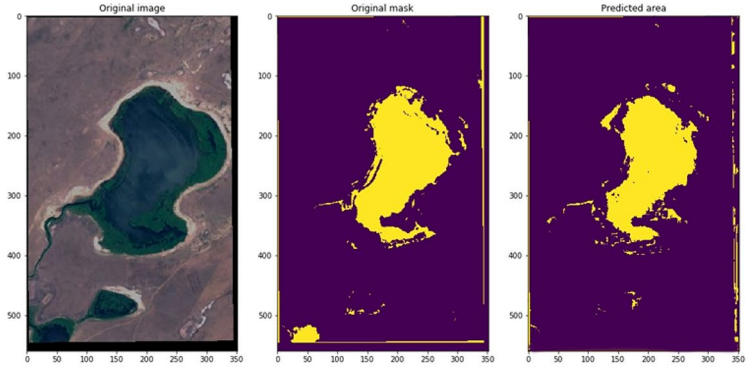
Table 3. Data set description.



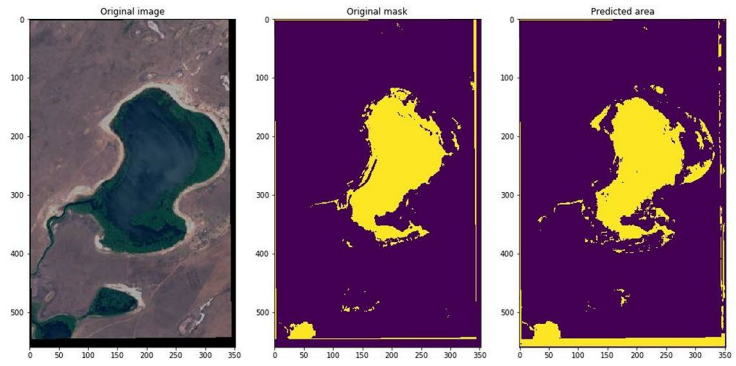
(a) SegNet with IoU 0.778



(b) ResNet: with IoU 0.754



(c) ENet with IoU 0.79



(d) Proposed Model with IoU 0.945

Fig. 4. Performance of the proposed and existing models

Methods	Precision	Recall	F1-Score	IoU
U-Net	0.831	0.83	0.832	0.834
U-Net + Enhanced Attention	0.89	0.88	0.894	0.892
U-Net + Residual	0.904	0.905	0.902	0.907
U-Net + Enhanced Attention + Residual	0.943	0.940	0.946	0.948

Table 4. Performance of the proposed model.

Methods	Trainable parameters weights (M)	Precision	Recall	F1-Score	IoU
U-Net $n = 3$(n-depth of U-Net)	8	0.824	0.87	0.823	0.827
U-Net $n = 4$(n-depth of U-Net)	8	0.826	0.828	0.829	0.829
U-Net $n = 5$(n-depth of U-Net)	9	0.832	0.835	0.829	0.832
U-Net + Enhanced Attention dilated convolution1	7	0.88	0.884	0.889	0.882
U-Net + Residual dilated convolution2	9	0.891	0.899	0.897	0.893
U-Net + Enhanced Attention + Residual	6	0.943	0.940	0.946	0.947

Table 5. Comparison with trainable parameters.

Network	Layers	Trainable Parameters	Running Time(ms)
SegNet[13]	21	30	282
ResNet[11]	22	19	145
FWE-Net[23]	23	14	187
Enet[14]	30	115	410
Deep Stacked Ensemble[16]	25	32	271
Proposed Method	22	27	219

Table 6. Comparison with different layers.

an 11th Gen Intel(R) Core (TM) i7-11700 processor (2.50 GHz), with 32GB of RAM and a 1 TB SSD, using Google Colab.

Ablation analysis

"Table 4 represents the performance of the proposed model", and "Table 5 represents a comparison with Trainable parameters". To demonstrate the enhanced efficacy of individual elements within the proposed AER-UNet for waterbody segmentation, ablation studies were conducted using Kaggle datasets. The results presented in the Tables illustrate the segmentation effectiveness in a sequential manner: starting from the U-Net, followed by U-Net with Enhanced attention, U-Net with residuals, and finally, the newly proposed AER U-Net (U-Net with both Enhanced Attention and residuals).

The results uncovered several valuable insights:

Utilizing the U-Net architecture alone yielded metrics of 0.831 for precision, 0.83 for recall, 0.832 for F1-Score, and 0.834 for IoU.

The integration of U-Net + Enhanced Attention connections led to significant improvements. Precision, recall, F1-Score, and IoU values rose to 0.89, 0.88, 0.894, and 0.892, respectively, with IoU reaching 0.907.

Further, incorporating U-Net + Enhanced Attention + Residual mechanisms enabled the model to identify crucial parameters while eliminating unnecessary ones. Consequently, precision, recall, F1-Score, and IoU metrics saw substantial improvements, reaching 0.943, 0.940, 0.946, and 0.948, respectively. "Table 6 represents the comparison with different Layers and Fig. 4 illustrates the visual representations of the proposed and existing model with IoU score".

Based on the outcomes, several valuable observations were made:

- For U-Net with a depth of $n = 3$, the precision, recall, F1-Score, and IoU metrics were 0.841, 0.87, 0.823, and 0.827, respectively. Slight variances were observed for $n = 4$ and $n = 5$. Considering parameters and complexity, opting for $n = 3$ is a preferable choice.
- The integration of U-Net with Enhanced Attention connections using dilated convolution1 and dilated convolution2 resulted in noteworthy improvements. Precision, recall, F1-Score, and IoU were calculated as 0.88, 0.884, 0.889, 0.882, and 0.891, 0.899, 0.897, 0.893, respectively.

Method	Precision	Recall	IoU
U-Net[9]	0.831	0.833	0.834
Block Attention-based U-Net[10]	0.811	0.813	0.815
ResNet[11]	0.750	0.752	0.754
DeepLabV3+[12]	0.719	0.721	0.723
SegNet[13]	0.771	0.773	0.778
ENet[14]	0.871	0.873	0.876
The Proposed Model	0.940	0.942	0.945

Table 7. Comparison of the proposed and existing models.

Introducing U-Net with with Enhanced Attention connections alongside Multi scale Resifual led to the identification of essential parameters while removing unnecessary ones. As a result, precision, recall, F1-Score, and IoU were enhanced to 0.943, 0.940, 0.946, and 0.947, respectively.

"Table 7 presents a comparison of the metrics between the proposed method" and several state-of-the-art models, including U-Net, ResNet, DeepLabV3, SegNet, and ENet. According to the statistics reported in Table 3, the modified U-Net model [30] demonstrates a higher IoU, a key metric for evaluating the segmentation performance of semantic images. This suggests that the proposed model outperforms the others overall in terms of segmentation accuracy. The main reasons behind the success of the proposed model:

1. By using attention layers, the model can focus on the most relevant features of the image (such as water bodies) and suppress less informative regions. This helps improve segmentation accuracy, especially in complex or cluttered images where distinguishing water from non-water regions is challenging [31].
2. Due to the multi-scale feature extraction, the model can accurately segment both large and small water bodies, capturing fine boundaries and irregular shapes that other models might miss.
3. ResNet introduces residual connections, which allow the network to learn more effectively by addressing the vanishing gradient problem and enabling deeper networks without losing performance. This leads to better feature extraction and improved segmentation accuracy[30–32].

Conclusion

This study successfully presents a robust and efficient deep learning approach for water body detection from satellite imagery, leveraging a modified U-Net architecture. The proposed model incorporates advanced features such as residual blocks, attention mechanisms, and dropout layers to improve segmentation accuracy and enhance generalizability. By employing a contracting-expanding path design with optimized activation functions, kernel initializers, and multi-channel feature maps, the model effectively captures and processes complex spatial features. Key architectural enhancements, including attention-refined skip connections and dropout regularization, address challenges like overfitting and the vanishing gradient problem. The use of Adam optimizers further accelerates computation and ensures effective training. Additionally, data preprocessing techniques such as resizing, scaling, and padding contribute to the model's precision and performance. This model excels even with small and hazy satellite images. Moreover, the AER U-Net architecture guarantees optimal results for water bodies located near land boundaries. The model also incorporates the Adam optimizer, which ensures faster computation times, making it an efficient solution for such tasks. The proposed approach achieves an IoU score of 0.94, demonstrating superior performance compared to existing methods. Its adaptability to high-resolution imagery and capability to accurately delineate water bodies make it a valuable tool for environmental monitoring, resource management, and disaster assessment.

Data availability

The datasets generated and/or analysed during the current study are available in the kaggle repository and real time dataset from [24], https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies.

Received: 27 December 2024; Accepted: 18 April 2025
Published online: 08 May 2025

References

1. World Water Assessment Programme (United Nations). *The United Nations World Water Development Report (No. 3)* (UNESCO Pub., 2009).
2. Drusch, M. et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **120**, 25–36. https://doi.org/10.1016/j.rse.2011.11.026 (2012).
3. McFeeters, S. K. The use of the normalized difference water index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **17**(7), 1425–1432. https://doi.org/10.1080/01431169608948714 (1996).
4. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **27**(14), 3025–3033. https://doi.org/10.1080/01431160600589179 (2006).
5. Feyisa, G. L., Meilby, H., Fensholt, R. & Proud, S. R. Automated water extraction index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **140**, 23–35. https://doi.org/10.1016/j.rse.2013.08.029 (2014).
6. Liao, A. et al. High-resolution remote sensing mapping of global land water. *Sci. China Earth Sci.* **57**, 2305–2316. https://doi.org/10.1007/s11430-014-4918-0 (2014).

7. Nath, R. K. & Deb, S. K. Water-body area extraction from high resolution satellite images-an introduction, review, and comparison. *Int. J. Image Process. (IJIP)* **3**(6), 265–384 (2010).
8. Zhu, X. X. et al. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sensing Mag.* **5**(4), 8–36. 10.1109/MGRS.2017.2762307 (2017).
9. Filatov, D., & Yar, G. N. A. H. Forest and water bodies segmentation through satellite images using u-net. arXiv preprint arXiv:2207.11222. https://doi.org/10.48550/arXiv.2207.11222 (2022).
10. Lwin, T. M. M. & Win, Z. M. Attention-based CNN model for semantic segmentation of remote sensing images. In 2024 5th International Conference on Advanced Information Technologies (ICAIT) (pp. 1–6). IEEE. (2024)., November https://doi.org/10.1109/ICAIT65209.2024.10754920
11. Silpalatha and jayadeva. Water bodies segmentation through satellite images using ResNet. *Int. J. Intell. Syst. Appl. Eng. IJISAE* **12**(4), 3253–3261 (2024).
12. Harika, A., Sivanpillai, R., Variyar, S. & Sowmya, V. V. V., Extracting water bodies in rgb images using deeplabv3 + algorithm. The international archives of the photogrammetry, remote sensing and spatial information sciences, **46**, 97–101. (2022). https://doi.org/10.5194/isprs-archives-XLVI-M-2-2022-97-2022, 2022.
13. Badrinarayanan, V. et al. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, **39**(12), 2481–2495. DOI: 10.1109/TPAMI.2016.2644615 (2017).
14. Paszke, S., Enet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint. (2016). https://doi.org/10.48550/arXiv.1606.02147
15. Akiyama, T. S., Junior, J. M., Gonçalves, W. N., de Araújo Carvalho, M. & Eltner, A. Evaluating different deep learning models for automatic water segmentation. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 4716–4719). IEEE. (2021)., July https://doi.org/10.1109/IGARSS47720.2021.9553345
16. Moradkhani, K. & Fathi, A. Segmentation of waterbodies in remote sensing images using deep stacked ensemble model. *Appl. Soft Comput.* **124**, 109038. https://doi.org/10.1016/j.asoc.2022.109038 (2022).
17. Yuan, K. et al. Deep-learning-based multispectral satellite image segmentation for water body detection. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **14**, 7422–7434. https://doi.org/10.1109/JSTARS.2021.3098678 (2021).
18. Feng, W., Sui, H., Huang, W., Xu, C. & An, K. Water body extraction from very high-resolution remote sensing imagery using deep U-Net and a superpixel-based conditional random field model. *IEEE Geosci. Remote Sens. Lett.* **16**(4), 618–622. https://doi.org/10.1109/LGRS.2018.2879492 (2018).
19. Zhang, X., Li, J. & Hua, Z. MRSE-Net: multiscale residuals and SE-attention network for water body segmentation from satellite images. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* **15**, 5049–5064. https://doi.org/10.1109/JSTARS.2022.3185245 (2022).
20. Li, K., Wang, J. & Yao, J. Effectiveness of machine learning methods for water segmentation with ROI as the label: A case study of the Tuul river in Mongolia. *Int. J. Appl. Earth Obs. Geoinf.* **103**, 102497. https://doi.org/10.1016/j.jag.2021.102497 (2021).
21. Chen, Y., Tang, L., Kan, Z., Bilal, M. & Li, Q. A novel water body extraction neural network (WBE-NN) for optical high-resolution multispectral imagery. *J. Hydrol.* **588**, 125092. https://doi.org/10.1016/j.jhydrol.2020.125092 (2020).
22. Miao, Z., Fu, K., Sun, H., Sun, X. & Yan, M. Automatic water-body segmentation from high-resolution satellite images via deep networks. *IEEE Geosci. Remote Sens. Lett.* **15**(4), 602–606. https://doi.org/10.1109/LGRS.2018.2794545 (2018).
23. Wang, J., Wang, S., Wang, F., Zhou, Y., Wang, Z., Ji, J., Zhao, Q. (2022). FWENet:a deep convolutional neural network for flood water body extraction based on SAR images. *Int. J. of Digital Earth*, 15(1), 345–361. https://doi.org/10.1080/17538947.2021.1995513.
24. https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies
25. Du, G., Cao, X., Liang, J., Chen, X. & Zhan, Y. Medical image segmentation based on U-net: A review. *J. Imaging Sci. Technol.* **64**(2). https://doi.org/10.2352/J.ImagingSci.Technol.2020.64.2.020508 (2020).
26. Reddy, K. R. & Dhuli, R. Detection of brain tumors from MR images using fuzzy thresholding and texture feature descriptor. *J. Supercomputing* **79**(8), 9288–9319. https://doi.org/10.1007/s11227-022-05033-x (2023).
27. Jonnala, N. S. & Gupta, N. SAR U-Net: Spatial attention residual U-Net structure for water body segmentation from remote sensing satellite images. *Multimedia Tools Appl.* **83**(15), 44425–44454. https://doi.org/10.1007/s11042-023-16965-8 (2024).
28. Jonnala, N. S., Gupta, N., Vasantrao, C. P. & Mishra, A. K. BCD-Unet: A novel water areas segmentation structure for remote sensing image. In 2023 7th international conference on intelligent computing and control systems (ICICCS) (pp. 1320–1325). IEEE. (2023)., May https://doi.org/10.1109/ICICCS56967.2023.10142694
29. Jonnala, N. S. & Gupta, N. NDR U-Net: segmentation of water bodies in remote sensing satellite images using nested dense residual U-Net. *Revue d'Intelligence Artificielle* **38**(3). https://doi.org/10.18280/ria.380324 (2024).
30. Vasantrao, C. P., Gupta, N., Jonnala, N. S. & Mishra, A. K. Dual adaptive model for change detection in multispectral images. In 2023 second international conference on electrical, electronics, information and communication technologies (ICEEICT) (pp. 1–6). IEEE. (2023)., April https://doi.org/10.1109/ICEEICT56924.2023.10156920
31. Chen, S., Liu, Y. & Zhang, C. Water-body segmentation for multi-spectral remote sensing images by feature pyramid enhancement and pixel pair matching. *Int. J. Remote Sens.* **42**(13), 5025–5043. https://doi.org/10.1080/01431161.2021.1906981 (2021).
32. Bansal, S. Long-wave bilayer graphene/hgcdte based GBp type-II superlattice unipolar barrier infrared detector. *Results Opt.* **12**, 100425. https://doi.org/10.1016/j.rio.2023.100425 (2023).
33. Bansal, S. Nature-inspired-based multi-objective hybrid algorithms to find near-OGRs for optical WDM systems and their comparison, Handbook of research on biomimicry in information retrieval and knowledge management, IGI global, 175–211. (2018). https://doi.org/10.4018/978-1-5225-3004-6.ch011
34. Jonnala, N. S. et al. DSIA U-Net: deep shallow interaction with attention mechanism UNet for remote sensing satellite images. *Sci. Rep.* **15**, 549. https://doi.org/10.1038/s41598-024-84134-4 (2025).

Acknowledgements

The authors acknowledge that the research Universiti Grant, Universiti Kebangsaan Malaysia, Geran Translasi: UKM-TR2024-10 conducting the research work. Moreover, this research acknowledges Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R10), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author contributions

N.S.J., S.S., Y.P. , P.C., B.P., made substantial contributions to design, analysis and characterization. S.B., P.U., K.P. participated in the conception, application and critical revision of the article for important intellectual content. M.R.I.F. and K.S.A.M. provided necessary instructions for analytical expression, data analysis for practical use and critical revision of the article purposes.

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.B., K.P. or M.R.I.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.