SSRG International Journal of Civil Engineering
ISSN: 2348-8352/ https://doi.org/10.14445/23488352/IJCE-V13I4P119

Volume 13 Issue 4, 318-332, April 2026

Original Article

A Deep Learning and Bayesian Regression Framework for Water Quality Assessment in Dam Reservoirs using Landsat Imagery

Sachchidanand Bhagat[1*], L.B. Roy[2]

[1,2]*Department of Civil Engineering, National Institute of Technology Patna (NIT Patna), Patna, Bihar, India.*

[1]*Corresponding Author : sachchidanandb.phd20.ce@nitp.ac.in*

Received: 14 January 2026 Revised: 16 February 2026 Accepted: 21 March 2026 Published: 28 April 2026

Abstract - *Water quality management in dam reservoirs plays a vital role in public health, aquatic ecosystems, and irrigation sustainability. Nagi and Nakti dam reservoirs in India, located in the state of Bihar, are affected by harmful substances such as pesticides, industrial chemicals, and untreated sewage water. These harmful substances increase nutrient levels such as nitrogen (N) and phosphorus (P), which affect aquatic life by increasing algal blooms, depleting oxygen, and altering the ratios of N: P. Hence, water quality is measured frequently in Nagi and Nakti dams. Traditional water quality management techniques, such as manual sampling, laboratory, and sensor-based methods, alter the contamination level due to delayed lab results and data loss because of human error, which prevents the predictive forecasting of eutrophication. To address the limitations of water quality management in dam reservoirs at an early stage, artificial intelligence-based techniques are proposed in this study. In this study, we proposed a Pixel-Based Optimized Deep Learning (PODL) framework, which consists of (I) deep Denoised Convolutional Neural Network (DnCNN) models to enhance the pixel level at the edges of Landsat satellite dam images to detect water and mass regions. Salp Swarm-Optimized layers of DnCNN (SSO-DnCNN) are proposed. The dual threshold graph cut segmentation technique is used to separate the mass region from the enhanced Landsat satellite image by minimizing energy at high green and red pixels. Then, textural features are extracted from the segmented mass region of the Landsat image. Bayesian-Optimized Multilinear Regression (BO-MLR) model predicts the water hydrogen level (pH) and dissolved oxygen (DO) from the extracted textural image and ground truth laboratory values, achieving an accuracy of about 91%-97%.*

Keywords - *Water quality monitoring, Landsat satellite imagery, Deep Convolutional Neural Networks, Bayesian multilevel regression, dam reservoir analysis.*

1. Introduction

Dam Reservoirs (DR) are used to store water using techniques, namely the dam linings and spillway design [1]. Water quality in DR is checked irregularly because of high maintenance costs and labour [2]. Optimal water quality maintenance is difficult due to water storage, flood mitigation, and hydropower generation in the reservoirs of Nagi and Nagathi in Bihar. The water quality is measured through factors such as physical, chemical, and biological factors [3, 4]. Plants absorb the chemical properties from water during the irrigation process, which reduces yields and affects public health [5].

Traditional water quality monitoring techniques, such as manual sampling, laboratory, and sensor-based methods, are expensive and time-consuming processes [6, 7]. The water quality is measured through the multispectral and hyperspectral images using deep learning [8, 9].

1.1. Problem Statement

Manual sample collection alters the contamination level and misses the contamination variation across the reservoir. Water moss disrupts the accurate prediction of DO and pH measurement levels.

1.2. Research Gap

The traditional machine learning model fails to capture the textural and spectral feature changes from satellite images. To improve the prediction accuracy of DO and pH levels, a Pixel-based Optimized Deep Learning Model (PODL) framework is proposed for continuous water quality management.

1.3. Contribution

The proposed PODL model predicts water quality by enhancing the edge and boundary pixels from mass and water areas of Landsat images and textural features with laboratory-

based features. For this study, the laboratory-tested water parameters and moss leaf samples were collected from the Nagi and Nakti reservoirs. A TxDyWT-based optimized DnCNN model is used to enhance the moss and water pixel features from Landsat images. A Bayesian-Optimized multilevel regression (BO- MLR) model is used to predict water quality from texture-level image features and laboratory values.

In this paper, Section 2 explains the literature review for monitoring water quality in different areas of the Dam Reservoir. Section 3 explains the methodology for measuring water quality in the Dam reservoir of Nagi and Nakti. Section 4 is the discussion section for the prediction of water quality from laboratory-based features and Landsat 8/9 image-based features.

2. Literature Survey

Table 1 shows the state-of-the-art techniques in water quality in specific geographic regions.

Table 1. State-of-the-Art Techniques in Water Quality Monitoring

Ref.	Study Area / Data Source	Methodology / Techniques	Parameters Estimated	Findings	Limitations / Gaps
[10]	Yeongju Dam, South Korea	3D numerical water quality model	General water quality variation	Reduce the water quality due to long residence time, gate operation changes, and low DO	limited predictive scalability
[11]	Gamcheon mid-basin, Korea	Load duration curves, drought analysis	River water quality indicators	Water quality changes due to cumulative rainy days, due to increased residence time, and higher nutrient concentration	Focused on river systems, not reservoirs
[12]	Freshwater systems	Drought-related water quality index	Composite water quality index	Low rainfall decreases DO and increases PH because of reduced inflow in the reservoir.	Not validated using satellite imagery
[13]	Edong, Gosam, Giheung reservoirs, South Korea	Sentinel-1 SAR, thresholding (Fixed, Otsu, Kittler–Illingworth)	Water body extent, storage	Water quality is measured through storage levels and hydrological behaviour	Not estimate chemical and biological features.
[14]	Hwanggang Dam, North Korea	Sentinel-2 imagery + DEM + hydrological model	Water level	Achieved $R^2 = 0.76$ for water level estimation	Restricted to water level; no quality parameters
[15]	El Guajaro reservoir, Colombia	Landsat-8 + multiple regression	Turbidity, DO, pH, depth, and conductivity	Measured through laboratory-based data	Regression-based; limited handling of noise and nonlinearity
[16]	Olushandja Dam, Namibia	Landsat-8 + regression models	Turbidity, DO, pH	Demonstrated feasibility of satellite-based retrieval	Performance is sensitive to atmospheric noise
[17]	Lake Dobczyce, Poland	Sentinel-2 + Lambert–Beer correction	Turbidity, chlorophyll-a	Produced spatial distribution maps of water quality	Limited to shallow reservoirs
[18]	Edong & Cheontae reservoirs, Korea	Multi-sensor fusion (SAR + optical) + K-means clustering	Water body delineation	Improved accuracy using NDWI–SAR fusion	No prediction of chemical parameters
[19]	Not specified	Hybrid DL with spatial–temporal features	Complex water quality parameters	Improved prediction accuracy	Model complexity and interpretability issues
[20]	Practical water management systems	Ensemble ML + visualization	Water quality indices	Enhanced real-time decision support	Requires extensive data infrastructure
[21]	Multi-sensor environments	Multimodal low-rank fusion with attention	Water quality parameters	Robust integration of heterogeneous data	Computationally intensive

[22]	Dynamic water systems	Sliding-window adaptive WQI	Water quality index	Improved responsiveness to rapid changes	Index-based, not parameter-specific
[23]	Not specified	Gated liquid neural networks	Water quality index	High stability and accuracy	Limited physical interpretability
	Proposed – Novelty	PODL framework	Water quality management	accuracy of about 91%-97%.	Sudden water quality fluctuations due to rain reduce the accuracy.

3. Proposed PODL Framework for Water Quality Management in Nagi and Nakti Dam Reservoir

The proposed PODL framework for water quality monitoring in the Nagi and Nakti Dam reservoir in Bihar, India, is shown in Figure 1. The captured Landsat images are pre-processed through the TxDyWT-based Optimized DnCNN model. The TxDyWT model separates the structural details of water and mass using low-frequency components. The Optimized DnCNN model enhances the edges through perspective projection on pixels and reduces the noise. Next, the double threshold graph cut segmentation model segments the mass from the water bodies by minimizing the energy in the ambiguous pixels. Finally, BO-MLR predicts the DO and pH levels from the dam reservoir Landsat image textural features and laboratory-based data. Therefore, the PODL framework continuously monitors the water quality of dam reservoirs.

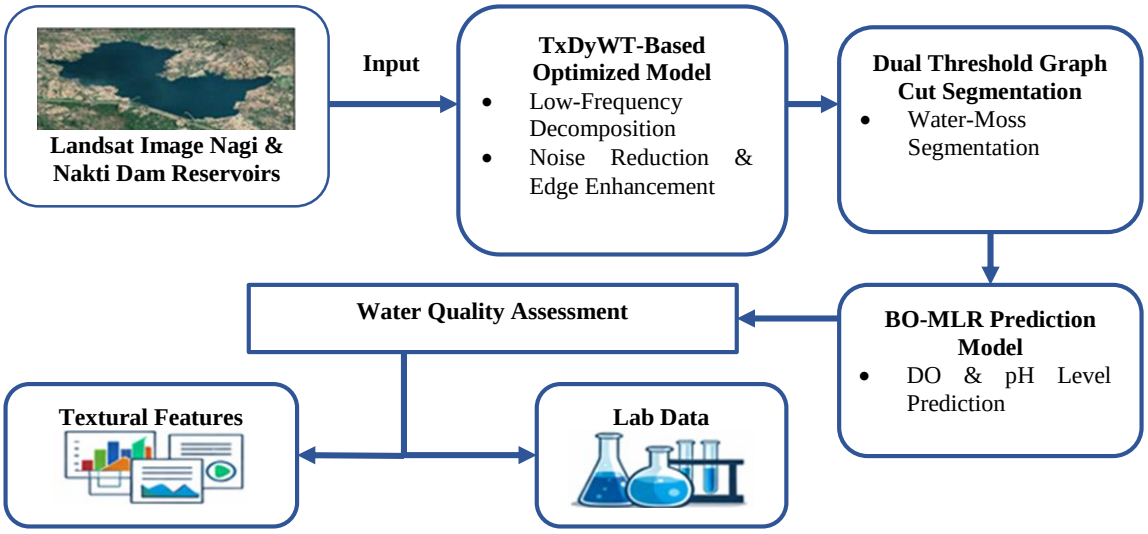


Fig. 1 Proposed PODL framework for water quality monitoring in the Bihar Dam Reservoir

3.1. Study Area and Dataset Collection

The data is collected from the Nagi and Nakti Dam Reservoir for continuous water quality management. Nagi Dam is located in Bihar with 425 ha of widespread area, 1,884 meters is the length of the DR, and 113.5 m is the height of the DR. DR storage capacity is 108 MCM. Nakti Dam was constructed in 1980 on the Nakti River. The Landsat images of Nagi and Nakti are obtained from January 2014 to December 2023, which show the water quality variation across the regions. All satellite images were downloaded from the USGS EarthExplorer platform. Image selection followed three main criteria: (i) cloud cover less than 10% to ensure minimal atmospheric interference, (ii) seasonal relevance, which is the seasonal variability in water quality, and (iii) availability of concurrent ground truth data from laboratory measurements for accurate model training and validation.

3.2. Landsat Band images Combinations – DR Water Quality Prediction (WQP)

Landsat band combinations, namely Band II, Band III, and Band IV, are used for change detection in DR for Water Quality Prediction (WQP) using the water body surface region enhancement.

The above Band combination enhances moss for better interpretation. In this paper, DR-WQP is performed using the Landsat band-II, band-III, and band-IV combination, which decreases the multi-collinearity and leads to the DR-WQP with more accuracy. Time series of Landsat band images are

used for DR-WQP. DR-WQP is performed for different periods of time and climatic conditions with more accuracy. Different band combinations of images at different climatic and environmental conditions, the DR-WQP changes are analyzed. The selection of bands of the Landsat images is performed to enhance the DR water quality monitoring. Figure 2 depicts the Landsat band. Landsat Band-2, Band-3, and Band-4 combinations were found to be suitable for water quality monitoring after analyzing different band combinations. Landsat images have high temporal resolution. Spatial resolution is 30 meters. Figure 2 shows the importance of band combination for the DR-WQP.

3.2.1. Preprocessing of Satellite Images

The raw Landsat pixel images are affected by sensor calibration, atmospheric scattering, and seasonal illumination changes.

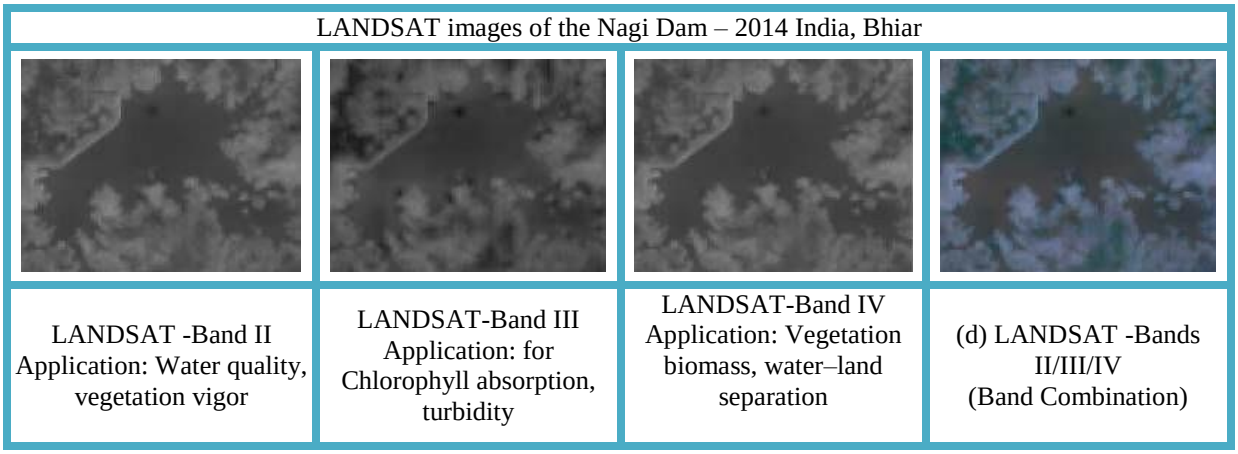


Fig. 2 Landsat bands applications for DR-WQP

Hence, the Landsat 8/9 images have non-random distortions and are affected by sensor artifacts. Hence, the radiometric calibration technique corrects the images of different dates and sensors by converting the image pixels into surface reflectance. Dark Object subtraction technique removes atmospheric scattering without altering the surface properties. Temporal normalization prevents the seasonal illumination by normalizing the pixel intensities among dominating features of Landsat images without altering true surface properties. Resizing enforces pixel-to-pixel correspondence, which is essential for temporal analysis and CNN processing. Temporal normalization further stabilizes pixel intensities across time, preventing seasonal illumination differences from dominating feature extraction. Figure 3 displays the process of pre-processing in the Landsat 8/9 image.

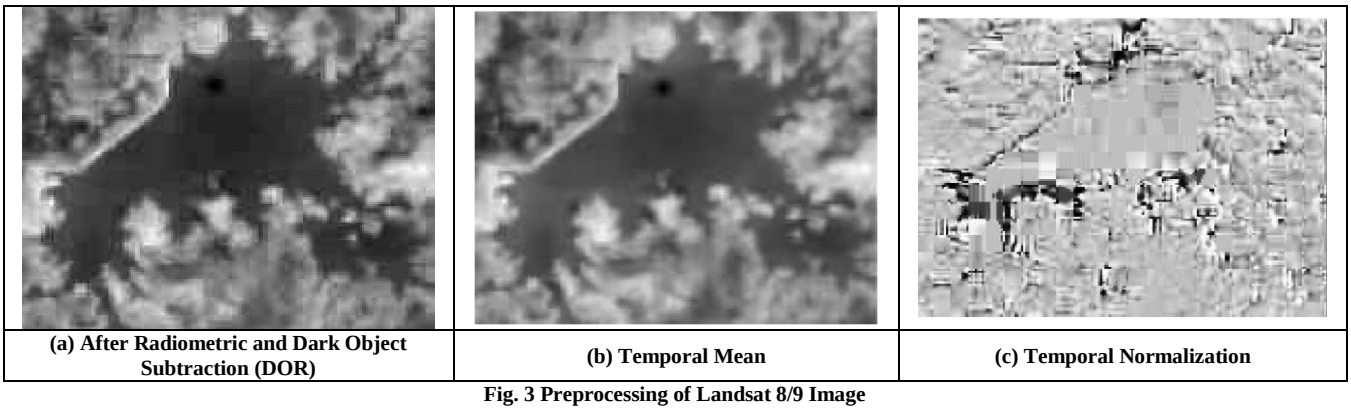


Fig. 3 Preprocessing of Landsat 8/9 Image

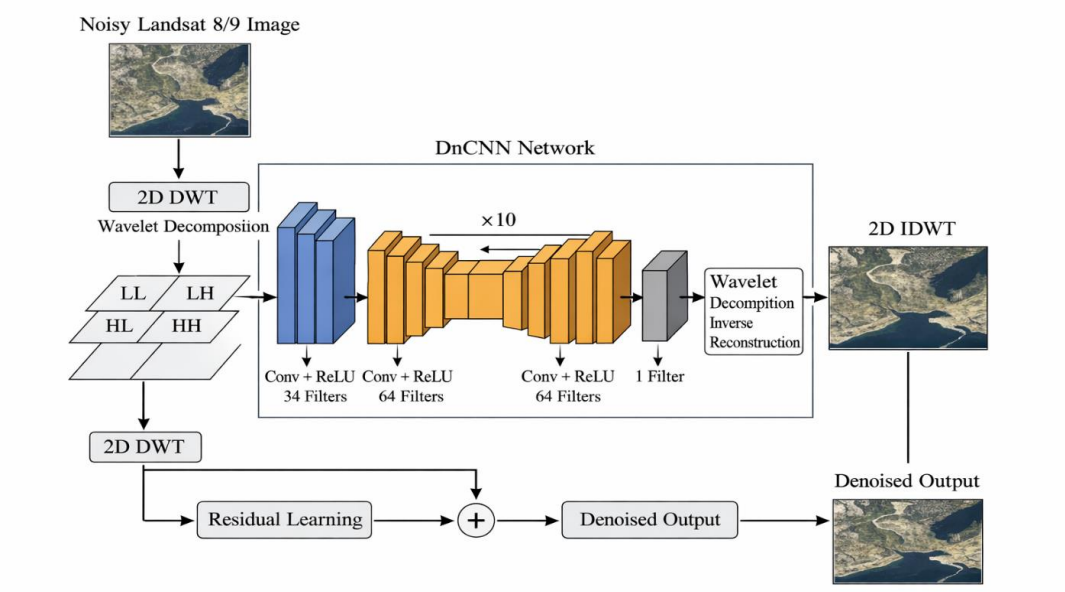
3.3. Landsat Dam Reservoir Image Enhancement Using Salp Swarm Optimized DnCNN TxDyWT - Perspective Projection

The proposed wavelet transformation-based optimized DnCNN model enhances the mass region from the water area of Landsat 8/9 images. Initially, the preprocessed Landsat 8/9 dam reservoir image is fed through TxDyWT. The TxDyWT decomposes the image into low-frequency and high-frequency components. TxDyWT separates the mass area from the shoreline and water by eliminating high-frequency components. To enhance the edges and boundaries of the mass region and to predict the water quality based on pH and

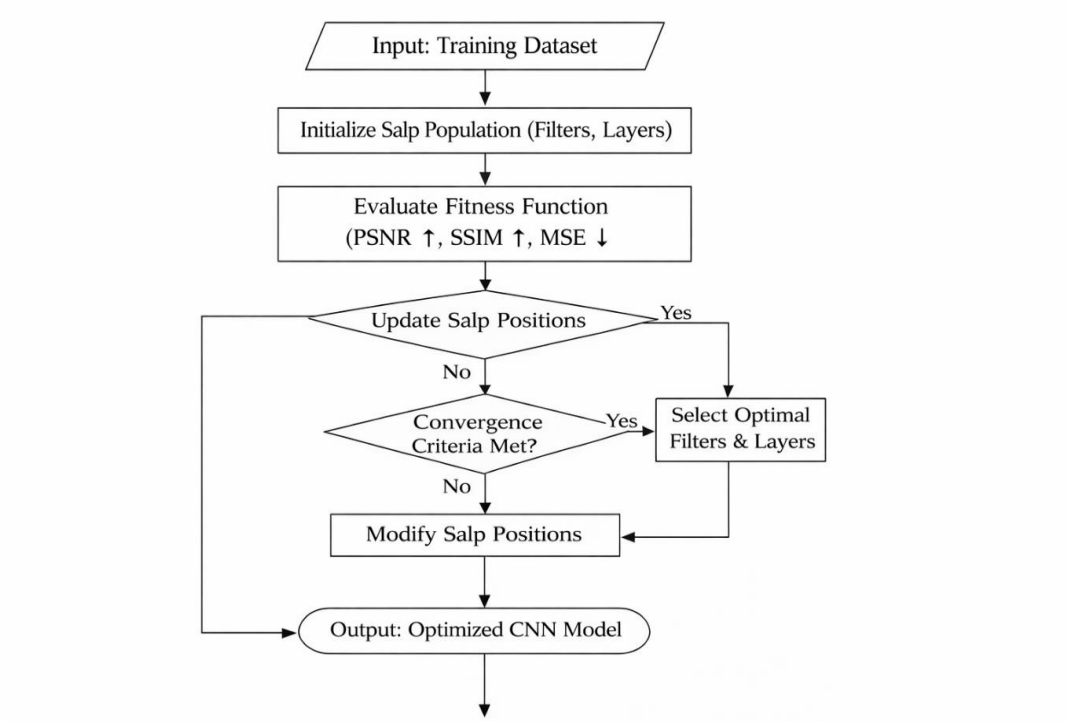
OD values, the optimized DnCNN model is used, which contains multiple convolutional layers and a ReLU layer to generate the enhanced image while preserving spatial consistency. Reservoirs have high spatial variability, such as water-land interfaces, vegetation, and shadows.

Table 2. Hyperparameters of DnCNN using different swarm optimization

Method	No. of Layers	Filters per Layer	Kernel Size	Activation Function	Learning Rate	Batch Size	Optimization Objective	Key Characteristics
Baseline DnCNN	17	64	3 × 3	ReLU	1e-3	64	MSE minimization	Fixed architecture, residual learning
Hyper-Parameter Optimized DnCNN	15–20 (adaptive)	32–64 (adaptive)	3 × 3	ReLU	1e-4 – 1e-3	32–64	PSNR / MSE	Grid/Random search, improved stability
PSO-DnCNN	14–18	48–64	3 × 3	ReLU	Optimized (≈1e-4)	64	PSNR maximization	Particle Swarm Optimization for filters & depth
RED-Ka DnCNN	16	64	3 × 3	ReLU + Skip	1e-4	64	Structural similarity (SSIM)	Recursive residual encoding, edge preservation
Fuzzy-DnCNN	15	64	3 × 3	Fuzzy-ReLU	1e-4	32	Noise uncertainty minimization	Fuzzy membership-based noise modeling
SSO-DnCNN (Salp Swarm)	12–16 (optimized)	32–64 (optimized)	3 × 3	ReLU	Auto-tuned	32–64	PSNR + SSIM (multi-objective)	Fast convergence, reduced overfitting, best trade-off



(a)



(b)

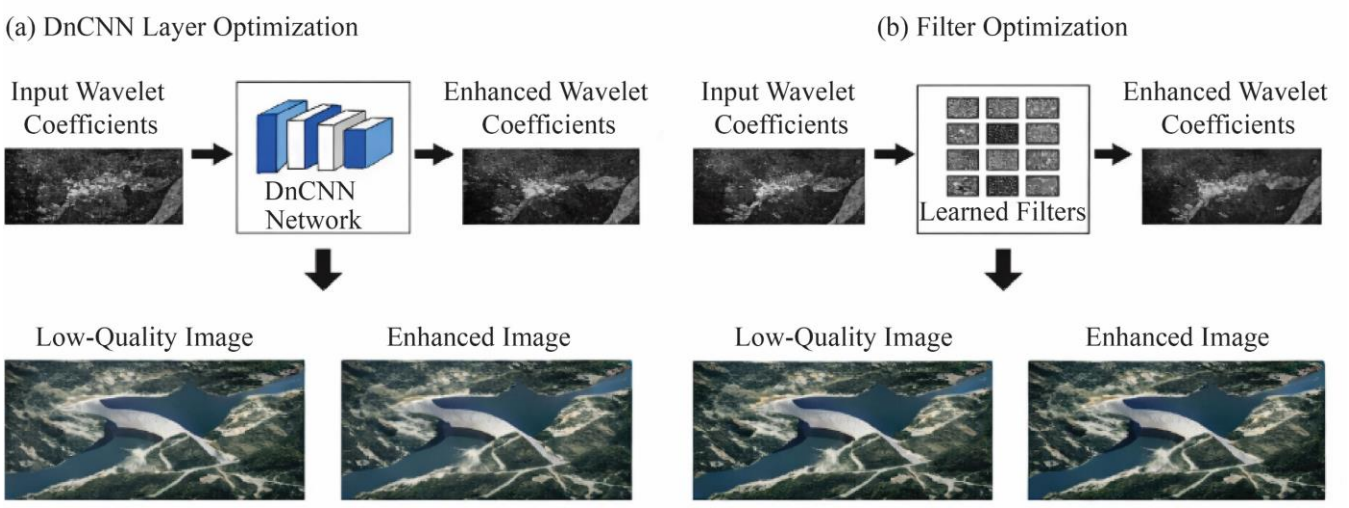


Fig. 4 Preprocessing steps in Landsat 8/9 Satellite image

A fixed DnCNN configuration fails to reduce the noise, overfits, and leads to a loss of fine structural details because of increasing network depth and filter count, which increases training time and the risk of overfitting. Therefore, to improve the quality of the Landsat image, the number of convolutional layers and filters is optimized through Landsat 8/9 images. Salp Swarm Optimization (SSO) is used to optimize the DnCNN parameters. Each candidate slap is represented as the depth of the convolutional layers and the filter size. The fitness function of each slap is defined as the SSIM and PSNR maximization while minimizing the MSE value. Leader and follower slaps update their positions by performing a balance between global exploration and local exploitation using equation (1).

Figure 4 shows the Landsat 8/9 satellite image enhancement process through optimized DnCNN with TxDyWT. Table 2 describes the hyperparameter optimization of convolutional layers and the number of filters used in the optimization process of the DnCNN network through different swarm-based optimization techniques.

$$\text{Fitness} = \alpha \times \text{PSNR} + \beta \times \text{SSIM} - \gamma \times \text{MSE} \quad (1)$$

In equation (1), α, β, and γ represent the weighting constants.

3.4. Dual Threshold Graph Cut (DGCT) Segmentation for Moss in Water

The double threshold-graph cut segmentation method is used to segment the Landsat 8/9 image into foreground and background images. DTGC segments the images based on two intensity thresholds: T1 and T2. T1 is the upper threshold, and T2 is the lower threshold. Landsat 8/9 image pixel intensities higher than the upper threshold (T1) are considered foreground seeds.

Image pixels with intensities lower than the lower threshold (T2) are considered background seeds. The remaining pixels are considered uncertain. The foreground and background seeds are used to construct the graph. Each pixel in the graph is connected to its neighboring pixels and is known as a node.

The energy minimization problem for segmenting the Landsat 8/9 dam reservoir is defined in Equation (2)

$$E(L) = \sum_{p \in \Omega} D_p(L_p) + \sum_{(p,q) \in N} V_{pq}(L_p, L_q) \quad (2)$$

In Equation (2), $E(L)$ represents the total cost of the segmented label. $\sum_{p \in \Omega} D_p(L_p)$ is the data fidelity term, which assigns the label L_p aligned to foreground and background pixels. $\sum_{(p,q) \in N} V_{pq}(L_p, L_q)$ is the smoothness constraint, which represents the pair of neighboring pixels and depends on the label changes between the adjacent intensity pixel differences among foreground and background for the uncertain pixel. Each uncertain pixel is aligned globally in either the background or foreground seeds. Hence, the energy minimization of the dual graph cut segmentation model distinguishes the water from the moss region in the Dam reservoir. Figure 5 shows the process of dual graph cut segmentation for separating moss from water in the Landsat 8/9 Dam reservoir image.

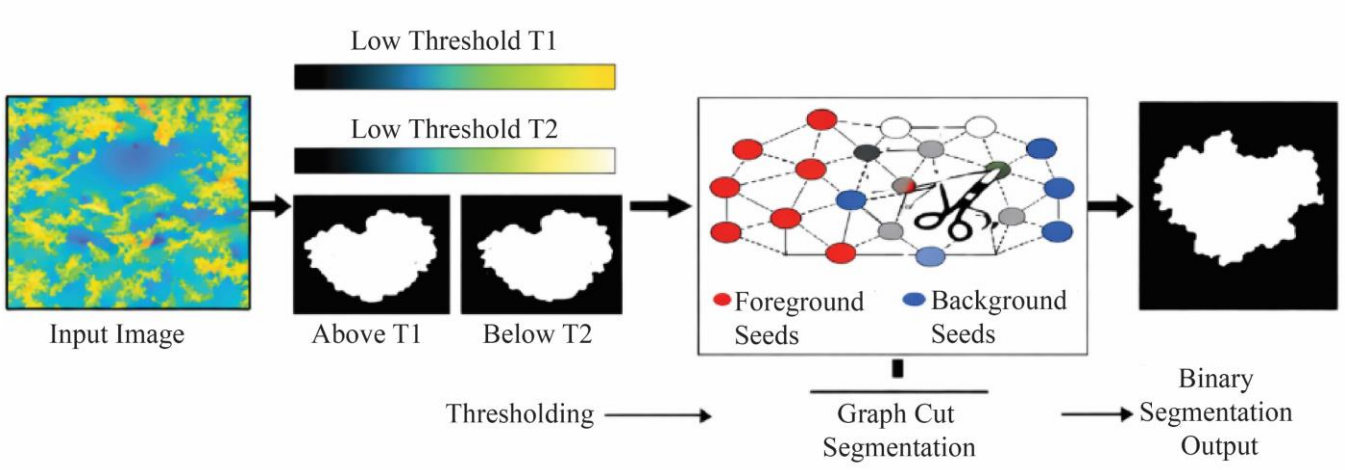


Fig. 5 DGCT technique to segment moss from water

3.5. DR-WQP in Naki and Nakathi

The water quality of the Naki and Nakti water dam is analyzed through metrics, as shown in Table 3. In Figure 6, Photograph Figure 6(a) is the Nagi, and Figure 6(b) is for the Nakti Dam water regions.

Nagi DR- India, Bihar	Nakti DR- India, Bihar
	

Fig. 6 Photo of Nagi and Nakti DR

Table 3. Water quality measurement methods

Water quality parameter Measurement method [Unit]	Nagi Dam range	Nakati Dam range
Centigrade mercury thermometer [°C]	24.5 – 30.2	25.1 – 31.0
Digital Nephelo-Turbidity Meter [NTU]	6.8 – 18.4	8.2 – 22.6
Digital pH meter —	7.2 – 8.4	7.0 – 8.2
Conductivity meter [µS/cm]	210 – 345	225 – 372
Conductivity-TDS meter [mg/L]	120 – 210	135 – 245
Winkler's modified method [mg/L]	5.1 – 7.4	4.6 – 6.9
Gravimetric method [mg/L]	18 – 46	22 – 58
Titrimetric method [mg/L]	110 – 185	125 – 205
EDTA titration [mg/L]	95 – 165	110 – 190
Argentometric method [mg/L]	18 – 42	22 – 55
5-day BOD test [mg/L]	1.8 – 3.6	2.2 – 4.1
Dichromate reflux method [mg/L]	14 – 28	18 – 34
UV spectrophotometric method [mg/L]	0.38 – 1.12	0.45 – 1.35
Ascorbic acid method [mg/L]	0.06 – 0.18	0.08 – 0.24
Turbidimetric method [mg/L]	12 – 34	15 – 41
Titrimetric method [mg/L]	3.5 – 9.6	4.2 – 11.8
Spectrophotometric method [µg/L]	8.4 – 22.7	10.6 – 28.9
Phenate method [mg/L]	0.12 – 0.46	0.18 – 0.62
ORP meter [mV]	+145 – +268	+128 – +245
Secchi disc [M]	0.85 – 1.65	0.72– 1.42

4. Results and Discussion

The proposed PODL framework is used to measure the water quality from the Nagi and Nakti Dam Reservoirs' Landsat 8/9 images. The pixel features are extracted from Landsat 8/9 images to predict water quality, and the BO-MLR technique is used to predict water quality using laboratory-based methods and Landsat 8/9 images. Figure 7 illustrates the TxDyWT used for the perspective projection of moss in DR. TxDyWT removes geometric distortions by separating high- and low-frequency components via multiresolution analysis. Then, the optimized DnCNN model preserves the edges and boundaries of moss from water by reducing the atmospheric and calibration noise level. The methods used for enhancing the moss in DR remove artifacts, as shown in Figure 8.

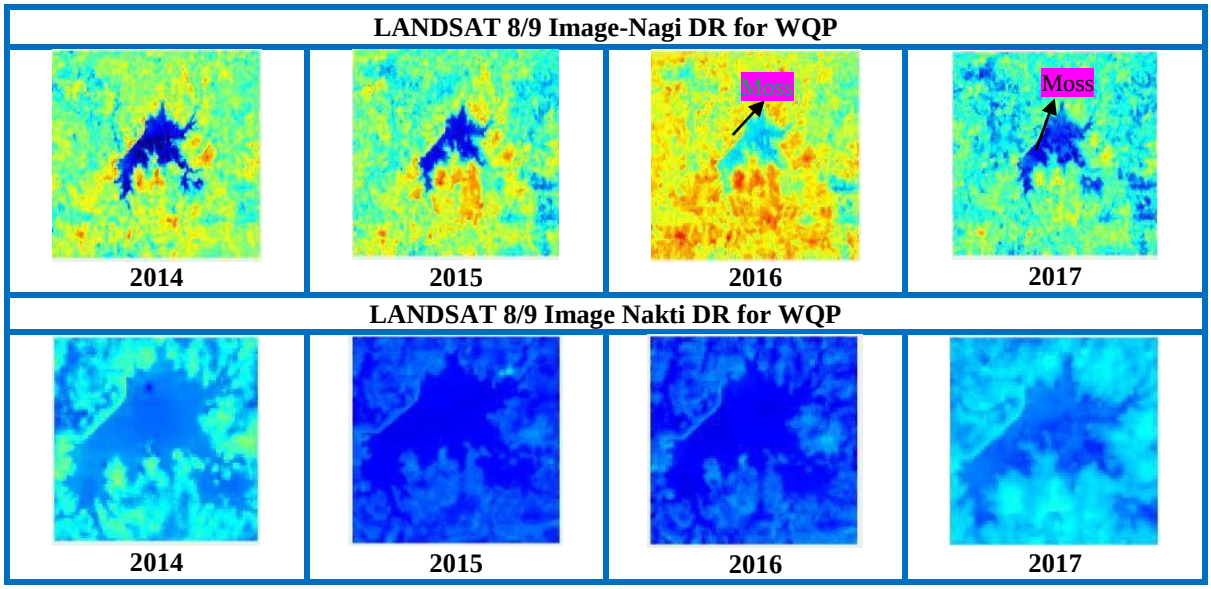


Fig. 7 Proposed TxTyDWT Algorithm for Nagi and Nakti Dam (2014-2017)

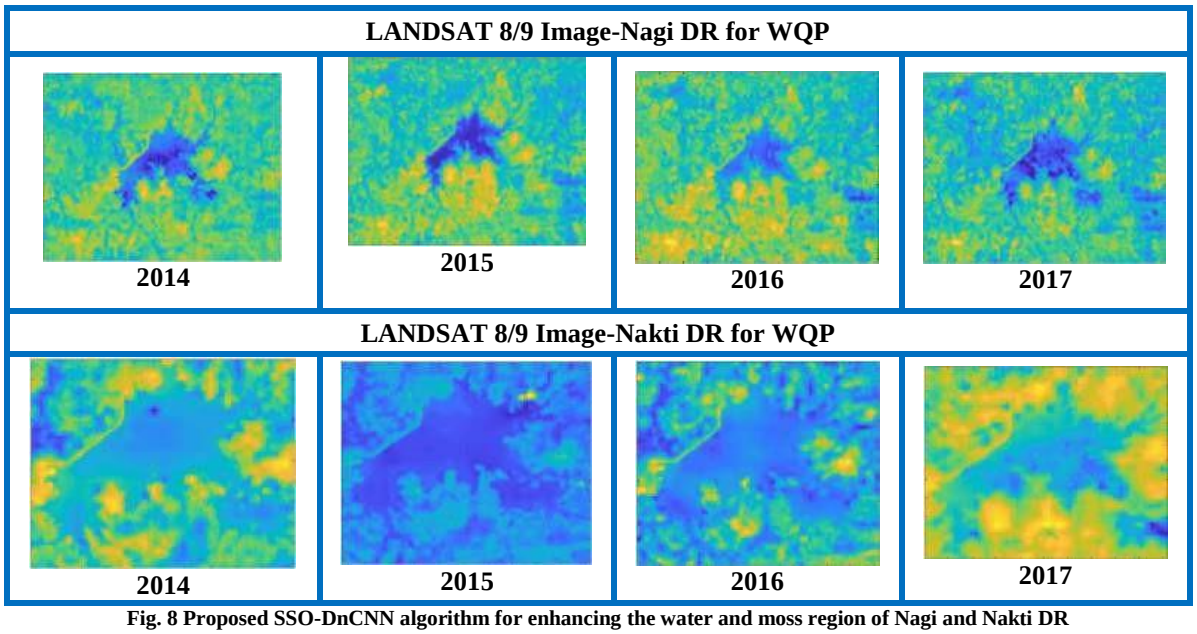


Fig. 8 Proposed SSO-DnCNN algorithm for enhancing the water and moss region of Nagi and Nakti DR

After enhancement, water and moss regions are segmented using the DTGC method. The graph is constructed based on dual thresholds, such as foreground and background seeds, for accurate detection of the moss region from the water surface. Figure 9 displays the separation of the water and moss region from Landsat 8/9 images.

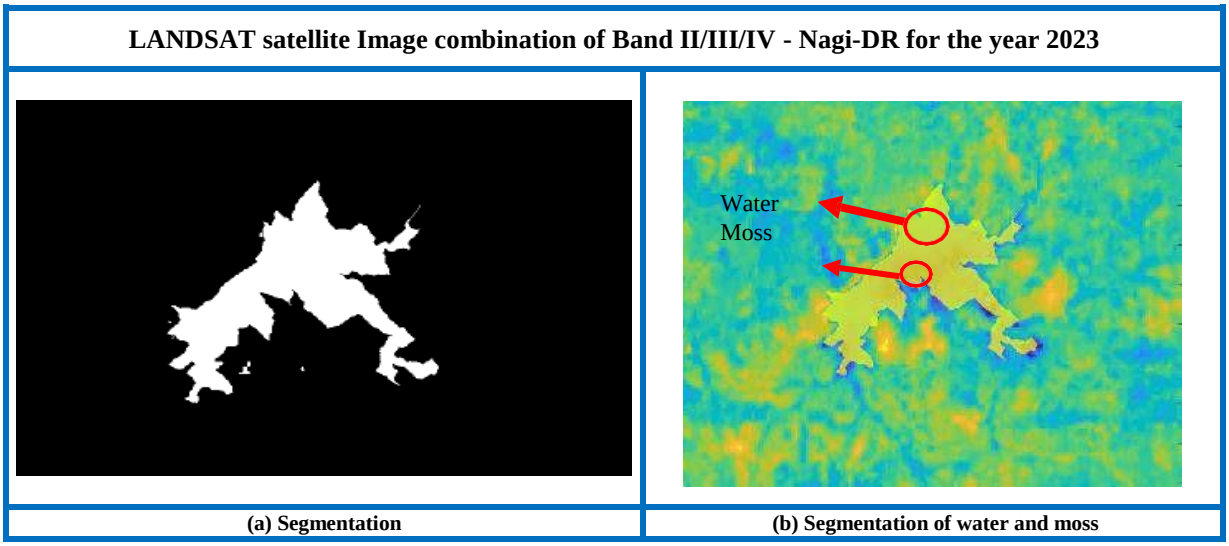


Fig. 9 Dual-Threshold Graph Cut (DTGC) algorithm segmented images for DR-WQP using the proposed PODL framework

The BO-MLR model is used to predict DO and pH values from physicochemical parameters and image-derived parameters such as conductivity, total dissolved solids, and turbidity. The BO-MLR model preserves the linear relationship between laboratory-based parameters and image-derived parameters. BO automatically minimizes the RMSE value by selecting the regularization value to predict the water quality measurements. Equations (3) and (4) represent the prediction of DO and pH values.

$$\widehat{DO} = \beta_0 + \beta_1 EC + \beta_2 TDS + \beta_3 Turbidity + \beta_4 Entropy_{water} + \beta_5 PSNR_{water} + \beta_6 SNR_{water} + \beta_7 Entropy_{moss} + \beta_8 WR + \beta_9 WG + \beta_{10} WB + \varepsilon_{DO} \quad (3)$$

$$\widehat{pH} = \alpha_0 + \alpha_1 EC + \alpha_2 TDS + \alpha_3 Turbidity + \alpha_4 PSNR_{moss} + \alpha_5 SNR_{moss} + \alpha_6 Mean_{water} + \alpha_7 Mean_{moss} + \alpha_8 MR + \alpha_9 MG + \alpha_{10} MB + \varepsilon_{pH} \quad (4)$$

The proposed framework was evaluated with standard statistical metrics. Their mathematical formulations are as follows in equations (5) to (10).

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (5)$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (6)$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|(4) \quad (7)$$

Peak Signal-to-Noise Ratio (PSNR)

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (8)$$

Signal-to-Noise Ratio (SNR)

$$SNR = 10\log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \quad (9)$$

Entropy (H)

$$H = -\sum_{i=1}^{k} p_i \log_2(p_i) \quad (10)$$

Table 4. Regression coefficient analysis of DO and its interpretation

Predictor	Coefficient (β)	Sign	Relative Importance	Interpretation
Turbidity	−0.312	−	Very High	Increased suspended particles reduce oxygen diffusion.
Entropy–Moss	−0.268	−	High	Dense moss increases biological oxygen consumption.
SNR–Water	+0.221	+	High	Clear water regions correlate with higher DO.
Conductivity	−0.184	−	Moderate	Higher ionic content reduces DO solubility.
TDS	−0.163	−	Moderate	Dissolved solids suppress oxygen availability.
PSNR–Water	+0.141	+	Low	Improved image clarity relates to stable DO.
WR	+0.097	+	Low	Reflects clean water spectral response.

Table 4 shows a regression coefficient analysis of dissolved oxygen (DO) using BO-MLR. Turbidity and entropy moss have high negative coefficients with reduced oxygen diffusion because of high turbidity and particle concentrations. The SNR of water has a strong positive value in water regions and is highly correlated with higher DO levels. Conductivity and Total Dissolved Solids (TDS) have moderate negative values, which reduce oxygen solubility. The positive values of PSNR water and WR, with their low positive coefficients, represent cleaner water and DO stability. Table 5 shows regression coefficients for pH prediction using BO-MLR. Conductivity has a strong positive coefficient, indicating that higher ionic concentrations in the dam reservoir are associated with increased alkalinity.

Table 5. Regression Coefficient Analysis of pH level and Its Interpretation

Predictor	Coefficient (α)	Sign	Relative Importance	Interpretation
Conductivity	+0.341	+	Very High	Higher ion concentration increases alkalinity.
TDS	+0.298	+	High	Dissolved salts influence buffering capacity.
WB (Blue channel – water)	+0.241	+	High	Indicates water clarity and pH stability
WG (Green channel – water)	+0.213	+	Moderate	Algal reflectance affects pH.
PSNR–Moss	−0.176	−	Moderate	Dense moss alters pH through respiration.
Mean–Water	+0.132	+	Low	Overall reflectance is linked to water chemistry.

TDS has a high positive coefficient, which shows that dissolved salts increase in the water, stabilizing the pH level. The positive coefficients of blue and green water channel intensities indicate that the presence of algae in the water regulates the pH level, due to dense moss growth altering the pH level through photosynthetic cycles. The smaller positive coefficient corresponds to mean water reflectance and is indirectly associated with overall water chemistry. Therefore,

the BO-MLR model appears to capture both chemical and biological controls on pH variation. Table 6 shows the prediction interval analysis of the BO-MLR model. Table 7 shows the seasonal performance analysis of DO and pH during the pre-monsoon and post-monsoon seasons. Table 8 shows the statistical significance between BO-MLR and standard MLR in achieving pH and DO prediction in water quality monitoring. Figure 10 shows the feature importance analysis of the BO-MLR model and identifies the predictors for both dissolved oxygen (DO) and pH estimation.

Table 6. Prediction Interval Analysis of DO and pH

Parameter	Mean Prediction	Mean Interval Width	95% PI
DO (mg/L)	9.82	±0.15	[9.67, 9.97]
pH	7.68	±0.08	[7.60, 7.76]

Table 7. Seasonal Performance Analysis of DO and pH and its interpretation

Season	R^2 (pH)	RMSE (pH)	R^2 (DO)	RMSE (DO)	Interpretation
Pre-Monsoon	0.972	0.152	0.969	0.158	High prediction accuracy due to increased turbidity, thermal stratification, and biological activity in low inflow conditions.
Post-Monsoon	0.978	0.138	0.975	0.145	Improved accuracy due to enhanced water mixing, dilution of pollutants, and stabilized physicochemical conditions.

Table 8. Statistical Significance between BO-MLR and MLR Model

Parameter	Test Used	Test Statistic	p-value	Significance	Interpretation
pH	Paired t-test	t = 4.21	0.0018	Significant	The BO-MLR model statistically significantly reduces prediction error and improves pH accuracy.
DO	Paired t-test	t = 4.67	0.0011	Significant	The higher test statistic and lower p-value indicate statistically significant improvement in dissolved oxygen prediction using BO-MLR.

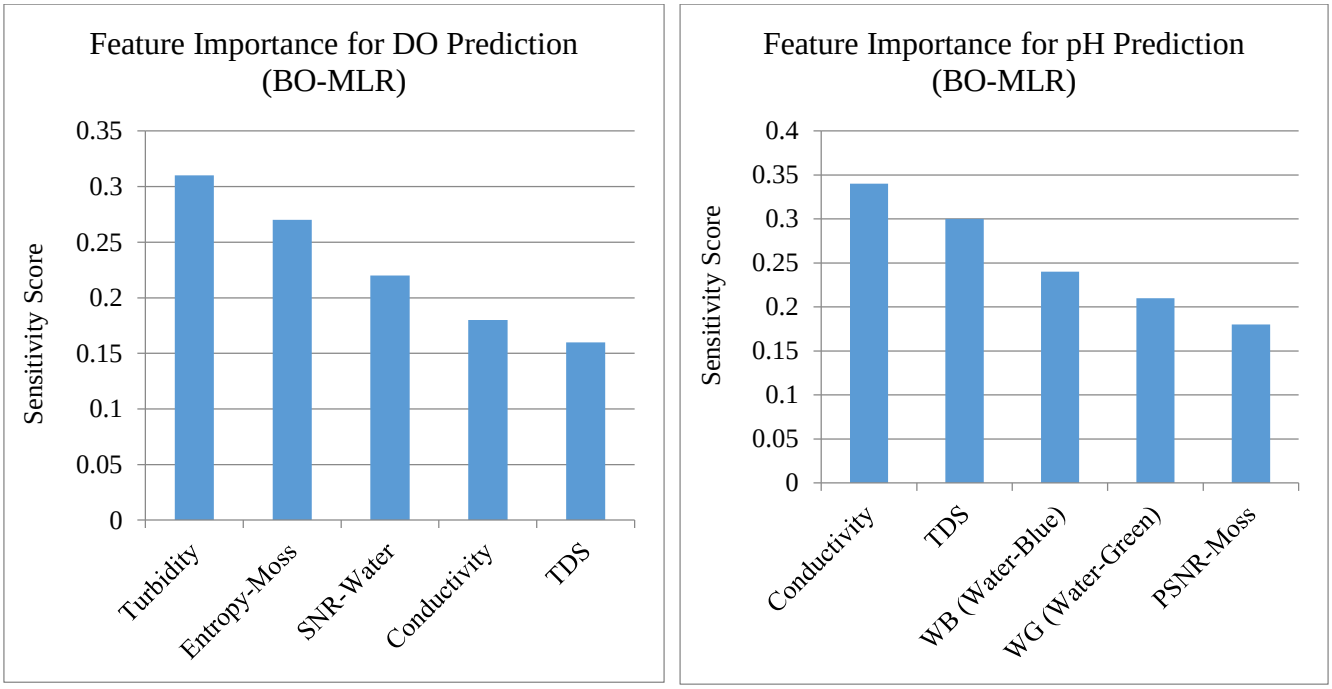


Fig. 10 Feature importance score in the prediction of DO and pH for DR-WQP using the proposed PODL framework

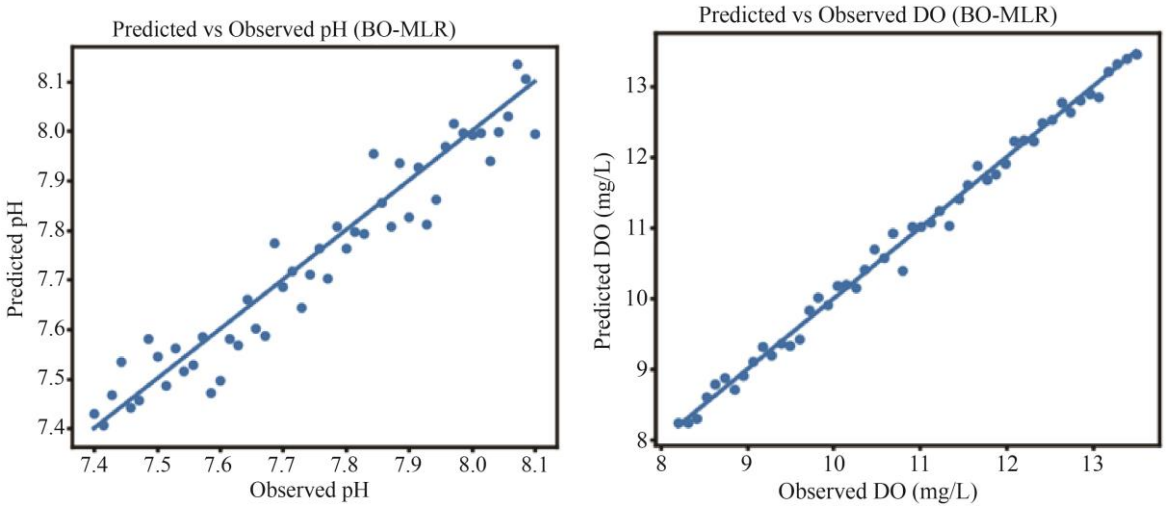


Fig. 11 Residual analysis of DO and pH for DR-WQP using the proposed PODL framework

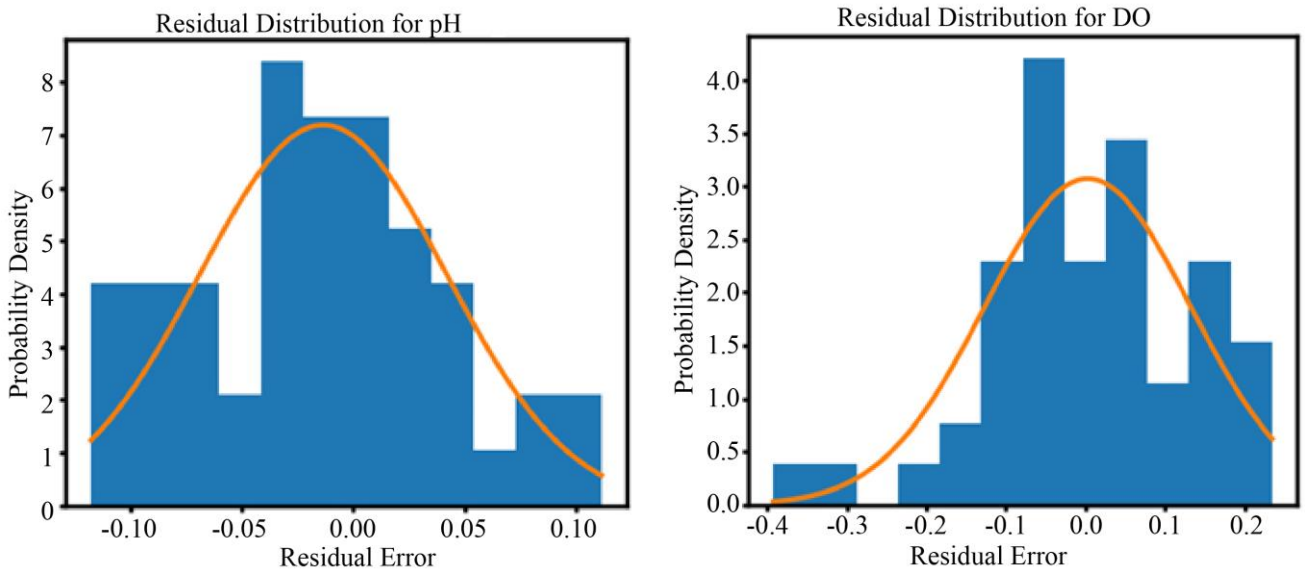


Fig. 12 Residual and prediction plots for DR-WQP using the proposed PODL framework

For DO prediction, turbidity is the most dominant feature, in which particles limit light penetration and oxygen diffusion. In pH prediction, conductivity and TDS are the most important features due to concentration and dissolved salts. The predicted versus observed plots are strongly correlated between the BO-MLR estimates and the measured values for both pH and Dissolved Oxygen (DO) in Figure 11. The plots show that the model accurately captures the relationship between predictors and the target variables for both pH and DO. The limited spread of residuals shows the model's stability and robustness in water. The residual distribution plots for both pH and Dissolved Oxygen (DO) show that the prediction errors are distributed around zero with a Gaussian shape and normal curve in Figure 12.

The narrow spread of residuals for pH shows low prediction and stable model behavior. The DO residuals are consistent with the higher variability of oxygen dynamics in aquatic systems. The prediction plots with 95% confidence intervals show the model's reliability in predicting pH and DO. For pH, the confidence intervals provide strong predictive stability and high certainty in estimation. For DO, the broader intervals reflect its greater temporal and spatial variability.

5. Discussions

The proposed PODL framework predicts the water quality of Nagi and Nakti Dam by analyzing Landsat 8/9 satellite imagery. The proposed PODL framework has SSO-DnCNN to reduce noise in the images and Dual Threshold Graph Cut Segmentation (DTGCS) to distinguish moss from water within the reservoirs. Furthermore, a BO-MLR model predicts water quality through analyzing pH and DO levels derived from both physiochemical properties and image-

based features such as color and texture. SSO-DnCNN shows stable performance across varying noise levels and pre- and post-monsoon seasonal conditions, providing high prediction accuracy compared to RKO-DnCNN, Fuzzy DnCNN, and PSO-DnCNN with a 95% confidence level. Prediction interval analysis of the PODL framework shows stable DO and pH prediction capabilities for quantifying water quality management. The PODL framework accurately predicts water quality based on DO and pH levels in both pre- and post-monsoon seasons. Although higher turbidity, thermal stratification, and reduced inflow degrade prediction accuracy during the pre-monsoon season, seasonal analysis and the framework's consistent performance across both seasonal conditions. The post-monsoon season improves the accuracy of water quality assessments through pollutant dilution and greater physical consistency in the model's behavior. Existing machine learning techniques provide lower R2 values and higher errors due to nonlinearity and noise. The proposed PODL framework achieves enhanced accuracy and robustness in measuring water quality within dam reservoirs. Table 9 displays the sensitivity analysis of the proposed PODL framework under different noise levels, and Table 10 compares the performance analysis of pH and DO across the proposed and existing algorithms.

Table 9. Sensitivity analysis of the proposed PODL framework under different noise levels

Noise Level (σ)	R^2 (mean ± std)	RMSE (mean ± std)	Quantified Change vs. Baseline	Interpretation
0.00 (Baseline)	0.976 ± 0.002	0.145 ± 0.006	Reference state	The model explains 97.6% of the variance with very low error, confirming strong predictive fidelity and stable learning.
0.01	0.973 ± 0.003	0.148 ± 0.007	R^2 ↓ 0.31%, RMSE ↑ 2.07%	Performance remains virtually unchanged, indicating high resistance to minor noise and preservation of learned patterns.
0.02	0.969 ± 0.004	0.152 ± 0.008	R^2 ↓ 0.72%, RMSE ↑ 4.83%	Slight degradation appears, but over 96.9% variance is still explained, reflecting continued robustness under realistic perturbations.
0.03	0.947 ± 0.006	0.159 ± 0.009	R^2 ↓ 2.97%, RMSE ↑ 9.66%	Noise begins to affect predictions meaningfully; explanatory power drops below 95%, and error growth accelerates.
0.05	0.922 ± 0.009	0.174 ± 0.010	R^2 ↓ 5.53%, RMSE ↑ 20.00%	The impact of noise becomes substantial, with reduced precision and weakened variance capture, though performance remains acceptable.

Table 10. Comparative performance metrics for pH and DO prediction across proposed and existing algorithms

Model	R^2 (pH)	RMSE (pH)	MAE (pH)	R^2 (DO)	RMSE (DO)	MAE (DO)	Interpretation
SSO-DnCNN + BO-MLR	0.964	0.163	0.128	0.961	0.169	0.132	Explains over 96% variance for both parameters with low and stable errors, indicating strong generalization and robustness to nonlinear noise.
RKO-DCNN + BO-MLR	0.976	0.145	0.112	0.973	0.152	0.118	Very high peak accuracy, but slightly less stable under perturbations compared to SSO-DnCNN in practical settings.
PSO-DCNN + BO-MLR	0.952	0.187	0.146	0.946	0.194	0.152	Moderate accuracy with noticeable error growth, suggesting limited convergence consistency.
Fuzzy-DCNN + BO-MLR	0.928	0.212	0.168	0.921	0.221	0.172	Acceptable variance capture, but higher errors indicate weaker feature discrimination.
[15]	0.902	0.241	0.189	0.896	0.249	0.193	Traditional approach with reduced nonlinear modeling capability.
[16]	0.887	0.258	0.201	0.872	0.263	0.207	Lower predictive reliability and increased deviation from true values.
[17]	0.861	0.274	0.215	0.849	0.279	0.219	Weakest performance, reflecting limited adaptability to complex water quality dynamics.

6. Limitations

Landsat 8/9 images have a spatial resolution of 30m, which is suitable for monitoring large-scale reservoirs. However, they are unable to capture small spatial heterogeneities such as shorelines, narrow channels, and localized pollutant inflow. Furthermore, the proposed framework monitors continuous water quality in the dam reservoir but is unable to detect sudden water quality fluctuations caused by heavy rainfall, accidental pollutant discharge, or increasing algal blooms.

7. Conclusion

The proposed PODL framework is used for continuous water-quality monitoring of the Nagi and Nakti DRs using Landsat 8/9 satellite imagery to predict pH and Dissolved Oxygen (DO) levels. The PODL model achieves 96% variance in both pH and DO, with low prediction errors. Through feature enhancement and noise suppression, the proposed PODL framework method for DR-WQP reduces the Root Mean Squared Error (RMSE) by 13-20% compared to existing techniques. The PODL framework achieved strong accuracy during the pre-monsoon period with high R2 values for both pH and DO. During the post-monsoon season, the model achieves R2 values of up to 0.978 for pH and 0.975 for DO due to enhanced mixing and pollutant dilution. In the future, multi-sensor fusion images such as Sentinel-2 optical and SAR imaging data will be used for continuous water quality monitoring.

Conflicts of Interest

The author(s) declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

The authors received no specific funding for this article.

Acknowledgments

Sachchidanand Bhagat: Conceptualization, Methodology, Software, Writing - original draft. L.B. Roy: Supervision, Review, Editing, Validation.

References

[1] Divakar Yadav et al., "Satellite Image Classification using Deep Learning Approach," *Earth Science Informatics*, vol. 17, no. 3, pp. 2495-2508, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Anja Batina, and Andrija Krtalić, "Integrating Remote Sensing Methods for Monitoring Lake Water Quality: A Comprehensive Review," *Hydrology*, vol. 11, no. 7, pp. 1-27, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Peng Zhang et al., "Forecasting Do of the River-Type Reservoirs using Input Variable Selection and Machine Learning Techniques-Taking Shuikou Reservoir in the Minjiang River as an Example," *Ecological Indicators*, vol. 155, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Muhammad Umar Farooq et al., "Assessment of Algorithm Performance on Predicting Total Dissolved Solids using Artificial Neural Network and Multiple Linear Regression for the Groundwater Data," *Water*, vol. 14, no. 13, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Benjamin J. Britton et al., "Evaluating Broadscale Deep Learning for Maya Settlement Detection in G-LiHT Lidar," *Journal of Archaeological Method and Theory*, vol. 33, no. 2, pp. 1-38, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] Amar Lokman, Wan Zakiah Wan Ismail, and Nor Azlina Ab Aziz, "Enhancing Water Quality Index Prediction Accuracy in Mranti Lake and Rivers in Malaysia using Regression Forest Model," *Applied Water Science*, vol. 16, no. 2, pp. 1-27, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] M. Uma Maheswari et al., "Efficient Drinking Water Quality Analysis using Machine Learning Model with Hyper-Parameter Tuning," *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, pp. 401-406, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Manish Kumar et al., "Quality Assessment and Monitoring of River Water using IoT Infrastructure," *IEEE Internet of Things Journal*, vol. 10, no. 12, pp. 10280-10290, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Swapna A. Jaywant, and Khalid Mahmood Arif, "Remote Sensing Techniques for Water Quality Monitoring: A Review," *Sensors*, vol. 24, no. 24, pp. 1-31, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Jin-Geun Park et al., "Contamination of Sediment after the Construction of the Yeongju Dam," *Journal of Environmental Science International*, vol. 33, no. 10, pp. 745-752, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Bugeon Jo et al., "Drought Impact on Water Quality Environment Through Linkage Analysis with Meteorological Data in the Gamcheon Mid-Basin," *Journal of Korea Water Resources Association*, vol. 56, no. 11, pp. 823-835, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Bu Li et al., "Enhancing Process-based Hydrological Models with Embedded Neural Networks: A Hybrid Approach," *Journal of Hydrology*, vol. 625, pp. 1-42, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Il-Hoon Choi, Eu-Ru Lee, and Hyung-Sup Jung, "Performance Comparison of Water Body Detection from Sentinel-1 SAR and Sentinel-2 Optical Imagery using Attention U-Net Model," *Korean Journal of Remote Sensing*, vol. 40, no. 5, pp. 507-523, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Sanae Kang, and Chul-Hee Lim, "Detecting Inaccessible Flood Damage in North Korea using Sentinel-1 Synthetic Aperture Radar Imagery: A Case of Flooding at 2023 Summer," *Korean Journal of Remote Sensing*, vol. 41, no. 1, pp. 173-184, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] F. Torres-Bejarano et al., "Impact of Land Cover Changes on Water Quality: An Application to the Guájaro Reservoir, Colombia," *International Journal of Environmental Science and Technology*, vol. 20, no. 4, pp. 3577-3590, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Alice Nureen Omondi et al., "Estimation and Mapping of Water Quality Parameters using Satellite Images: A Case Study of Two Rivers Dam, Kenya," *Water Practice and Technology*, vol. 18, no. 2, pp. 428-443, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[17] Nidal M. Hussein, and Mohammed N. Assaf, "Multispectral Remote Sensing Utilization for Monitoring Chlorophyll-A Levels in Inland Water Bodies in Jordan," *The Scientific World Journal*, vol. 2020, no. 1, pp. 1-14, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Liming Zhang et al., "An Infrared and Visible Image Fusion Algorithm based on Resnet-152," *Multimedia Tools and Applications*, vol. 81, no. 7, pp. 9277-9287, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Jie Long et al., "Application of an Improved LSTM Model based on FECA and CEEMDAN-VMD Decomposition in Water Quality Prediction," *Scientific Reports*, vol. 15, no. 1, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[20] Yas Barzegar et al., "Data-Centric Water Safety Monitoring: A Machine Learning Pipeline with Intelligent Feature Selection for Potability Prediction," *Procedia Computer Science*, vol. 270, pp. 6035-6044, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Jing Bi et al., "STMF: A Spatiotemporal Multimodal Fusion Model for Long-Term Water Quality Forecasting," *IEEE Internet of Things Journal*, vol. 12, no. 18, pp. 37146-37159, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Bhavana Reddy Tadimarri, and Ahyoung Lee, "IoT-based Real-Time Water Quality Assessment with Predictive Analytics," *2025 IEEE Opportunity Research Scholars Symposium (ORSS)*, Atlanta, GA, USA, pp. 1-4, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] R. Maruthamuthu et al., "Enhanced Water Quality Index Prediction using a Dual-Key Convolutional Transformer Autoencoder Network Integrating with Poplar Optimization," *Franklin Open*, vol. 13, pp. 1-13, 2025. [CrossRef] [Google Scholar] [Publisher Link]