Global NEST Journal, **28** (8), 08643
Available online: 16/09/2026

A Coupled RF-GWO-CNN Model for Predicting Water Gross Alpha Radioactivity Using Routine Water Quality Indicators

Xuemei Wu[1,2], Zongxi He[3,4†], Gengming Zhou[1,2], Haijun Zhang[1,2], Bo Du[1,2], Suhua Wang[1,2], Qiao Liu[1,2], Zhaoxu Lu[1,2], Jing Tian[1,2*]

[1]Nanchong Ecological and Environmental Monitoring Central Station of Sichuan Province, Nanchong, Sichuan, 637000, China
[2]Intelligent Recognition and Precise Control of Emerging Contaminants Key Laboratory of Nanchong, Nanchong, Sichuan, 637000, China
[3]Department of Nuclear Medicine, Affiliated Hospital of North Sichuan Medical College, Nanchong, Sichuan, 637000, China
[4]Nuclear Medicine and Radiation Safety Key Laboratory of Sichuan Province, North Sichuan Medical College, Nanchong, Sichuan, 637000, China

*Correspondence e-mail: ncjcz123dhd@163.com
[†]These authors (Xuemei Wu and Zongxi He) contributed equally to this work and should be considered co-first authors.

Abstract

Monitoring gross alpha radioactivity in water is essential for assessing the health risks associated with nuclear energy utilization. In the Jialing River basin (Nanchong), prediction is hindered by scarce samples, discontinuous monitoring, and the statistical fluctuations of low-background radiation. This paper proposes an integrated RF-GWO-CNN framework that couples Random Forest (RF) feature selection with a Convolutional Neural Network (CNN) tuned by Gray Wolf Optimization (GWO). The RF Gini importance ranking narrows fourteen candidate inputs (eight routine water-quality parameters plus geographic, meteorological and temporal variables) to nine key predictors, and GWO automatically optimizes the learning rate, training epochs and regularization of the CNN. Using 100 samples from 15 monitoring sites, the framework is compared with backpropagation (BP), support vector machine (SVM), a standard CNN, and two tree-ensemble regressors under repeated five-fold cross-validation. It reaches a cross-validated RMSE of 0.017 Bq/L and MAE of 0.013 Bq/L, on par with the random-forest and CNN baselines and significantly better than SVM (Wilcoxon signed-rank test, $p < 0.05$). The strength of the approach is an interpretable, automatically tuned pipeline that turns routine, rapidly measured indicators into early warning of gross alpha radioactivity where monitoring data are sparse.

Keywords: Environmental radiation early warning; Low-background radiation; Feature importance ranking; Random Forest; Gray Wolf Optimization; Metaheuristic hyperparameter tuning; Data-scarce basin.

OPEN ACCESS

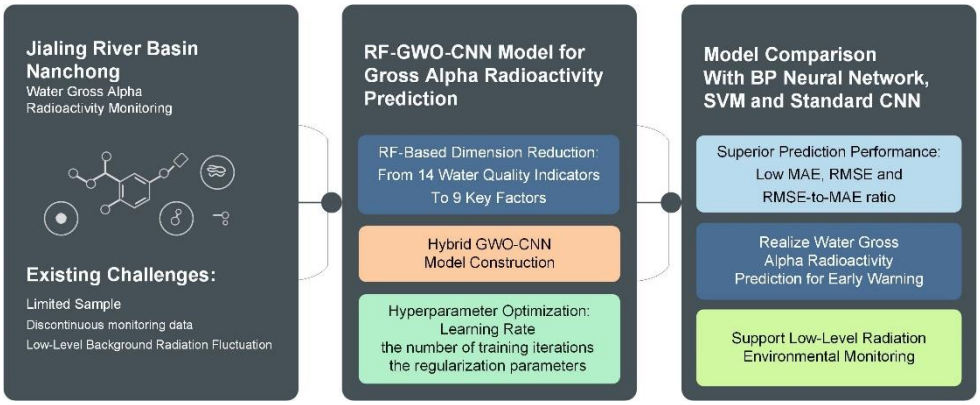
Received: 18/05/2024,

Accepted: 22/08/2026,

Available online: 16/09/2026

Copyright: © 2026 The Author(s). Published in Global NEST Journal. This article is licensed under the terms and conditions of the Creative Commons Attribution 4.0 International license (CC BY 4.0).

Graphical abstract



1. Introduction

In recent years, nuclear energy has been increasingly used in the energy sector. Its environmental risks have also drawn more attention (Klingelhöfer *et al.*, 2024; Ghisellini *et al.*, 2025). As a result, the radioactive safety of aquatic ecosystems is now recognized as a global environmental concern, especially during emergencies. The 2011 Fukushima nuclear accident in Japan is a clear example. The subsequent decision to release contaminated water into the ocean has raised international concerns. These concerns include the spread of marine radioactive pollution, long-term ecological impacts, and potential health risks (IAEA, 2022). Nanchong City is in the middle reaches of the Jialing River. Its water quality affects the drinking water safety of more than five million residents (Yuanping Zhuo, 2018). It is also considered important for the ecological security of the lower Jialing River and the Yangtze River basin (He *et al.*, 2023). Therefore, establishing a scientific and systematic radiation early warning and prevention mechanism is of great significance for ensuring the safety of the regional water environment.

Nanchong is a city with a low level of industrialization in Sichuan Province (W. Yang *et al.*, 2019). It currently has no automated monitoring stations for water radioactivity. Therefore, predictive studies on water radioactivity in this area face several challenges. First, the gross alpha radioactivity in environmental water is usually at a low background level. Its statistical fluctuations also bring inherent uncertainty to the data. Second, manual monitoring is a cumbersome and time-consuming process , with only a limited amount of data provided. As a result, it cannot accurately capture continuous changes in radioactivity levels. In addition, there are no sample data from sudden radioactive incidents. This makes it difficult to build traditional prediction models that rely on large amounts of accident data. Third, radioactivity levels may be related to some routine water quality indicators and geospatial data. But these relationships are complex and not yet fully understood. Previous studies show that gross alpha radioactivity is statistically correlated with conductivity, nitrate, and manganese (Al-Amir *et al.*, 2012; Ali, Matar, *et al.*, 2025; Tăban *et al.*, 2023). It is also influenced by geo-hydrological factors, such as water source type and water-rock interactions. Therefore, it is a practical idea to use routine water quality indicators for prediction. These indicators are easy to obtain and less time-consuming. This approach is technically feasible and has potential for real-world application.

Artificial intelligence (AI) and machine learning (ML) technologies excel at capturing nonlinear relationships within data, thereby offering robust tools for addressing prediction challenges in complex environmental systems. Numerous studies have demonstrated their efficacy in comprehensive water quality assessment and parameter prediction (Chen *et al.*, 2020; Sathya, P. *et al.*, 2024; Visu, P. *et al.*, 2024; Zhu *et al.*, 2023). Among these, backpropagation (BP) neural networks possess self-learning and adaptive capabilities, fault tolerance, and strong nonlinear mapping abilities, enabling them to effectively approximate arbitrary continuous nonlinear functions (Defeng Zhang, 2021). As one of the earliest models applied to environmental water quality prediction (Maier & Dandy, 2000), BP neural networks have been widely adopted. For instance, Luo *et al.* (2025) employed a BP neural network to analyze five surface water parameters–including turbidity, total nitrogen (TN) concentration, and grayscale values–in the middle reaches of the Yangtze River basin. Their model achieved an R^2 of 0.798 with a mean absolute percentage error (MAPE) of 29.0% for turbidity, and an R^2 of 0.536 with a MAPE of 26.3% for TN (Luo *et al.*, 2025). These results confirm the feasibility of BP neural networks for water quality prediction. However, constrained by factors such as training sample size and input variable complexity, BP neural networks exhibit inherent limitations, including a tendency to fall into local minima, unstable training processes, and overfitting. These issues restrict their generalization capability and reliability in practical applications.

In contrast, support vector machines (SVM) construct optimal classification hyperplanes in high-dimensional feature spaces based on the principle of structural risk

minimization. In water quality prediction, SVM often delivers more stable and reliable predictive accuracy than neural networks, particularly when dealing with limited data. Abobakr S. A. Y. *et al.* (2019) developed an efficient SVM model to predict water quality in the Langat River basin. Their model demonstrated accurate simulation capabilities for ungauged basins or those lacking sufficient monitoring stations, achieving maximum R^2 values of 0.998 and 0.979, with mean squared error (MSE) values ranging from 0.004 to 0.681 (Abobakr Yahya *et al.*, 2019). Nevertheless, the predictive performance of SVM is highly dependent on the selection of kernel functions and penalty parameters, making its effectiveness susceptible to parameter tuning (Xiaochuan Wang *et al.*, 2013).

Convolutional neural networks (CNNs) automatically extract local spatial features through convolutional kernels and achieve feature dimension reduction and aggregation via pooling operations, demonstrating exceptional performance when processing spatially correlated water quality monitoring data (A. Helaly *et al.*, 2025; Muñoz-Alegría *et al.*, 2025; S. Yang *et al.*, 2024). However, hyperparameter tuning for deep models such as CNNs typically relies on empirical methods and incurs high computational costs (Ladi *et al.*, 2022), which limits their efficiency in practical applications.

To address the above issues, with the aim of predicting the gross alpha radioactivity based on routine water quality indicators, this study selected the mainstream of the Jialing River, important lakes and reservoirs, and typical small watersheds in Nanchong City, Sichuan Province as the research objects. The following work was carried out:(1) To identify potential correlations between routine water quality indicators and gross alpha radioactivity, an RF algorithm was employed to evaluate feature importance. Subsequently, a standard convolutional neural network (CNN) prediction framework was used to select the optimal feature subset. (2) Based on this optimal subset, the geochemical processes linking input and output features were analyzed for correlation. (3) Considering the modeling characteristics under specific non-accident low-background data conditions, a CNN prediction model optimized by the Gray Wolf Optimization (GWO) algorithm was proposed, leveraging the feature correlations between routine water quality indicators and gross alpha radioactivity. This approach enables rapid, automatic, and global optimization of key hyperparameters, thereby improving prediction accuracy while reducing computational time.

2. Materials and Methods

In low-level environmental water, gross alpha radioactivity has complex nonlinear relationships with routine water quality indicators (Al-Amir *et al.*, 2012; Ali, Matar, *et al.*, 2025; Tăban *et al.*, 2023). It is also affected by statistical fluctuations. Because of these factors, conventional models have difficulty making accurate predictions. Although CNNs show promise for prediction tasks, their accuracy remains sensitive to hyperparameter settings, making optimal hyperparameter configuration a significant challenge. To solve this problem, a hybrid prediction framework is proposed in this study. It combines an RF for feature selection with the GWO algorithm for CNN hyperparameter optimization. First, the RF algorithm is used to assess the importance of each water quality indicator. This helps reduce the number of input features. Next, the GWO algorithm is introduced. It uses its strong global search ability to automatically optimize the key hyperparameters of CNN. Studies show that GWO has clear advantages for CNN hyperparameter optimization. Optimized models perform better. They also cut training time by more than 11% (Kaissar *et al.*, 2025).In water quality prediction, this optimization method saves time compared to traditional tuning. It also avoids unnecessary searches (Hou *et al.*, 2025). The modeling process is shown in **Figure 1.**

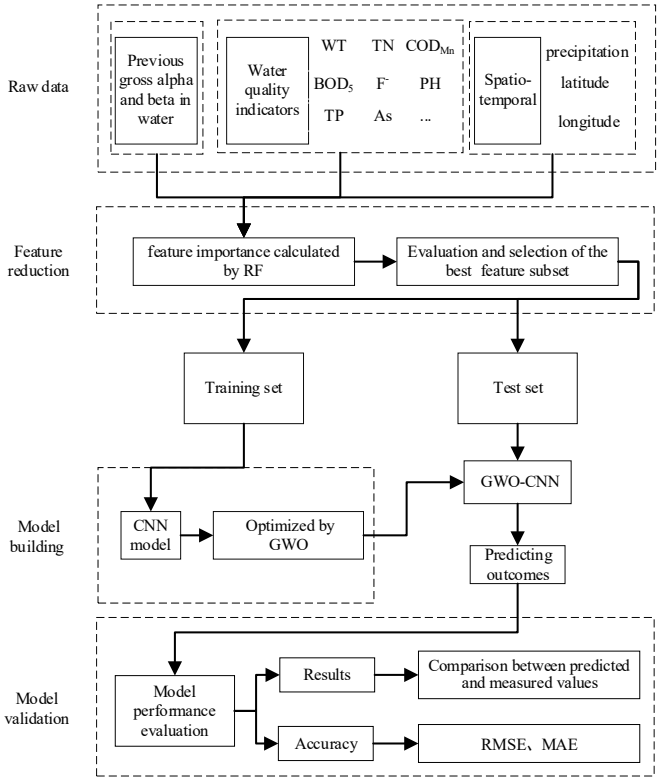


Figure 1. Algorithm flowchart of the RF-GWO-CNN model.

2.1. *Data Sources*

This study uses routine water quality indicators to predict gross alpha radioactivity in water. Data were collected from the Nanchong Ecological and Environmental Monitoring Central Station in Sichuan Province. Sampling was done at 15 sites (**Figure 2**). These sites covered the main stem of the Jialing River, major lakes and reservoirs, and small watersheds. From 2019 to 2024, eight routine water quality parameters were measured at each site. These were water temperature (WT), pH, permanganate index (COD_{Mn}), biochemical oxygen demand (BOD_5), Fluorine reagents (F^-), total phosphorus (TP), TN, and arsenic (As). Gross alpha and gross beta radioactivity were also measured at the same locations. Sample collection, analysis, and quality control were carried out in accordance with national standards. **Table 1** shows the abbreviations, units, and methods used for each parameter. Geographic and meteorological data were also collected for each site,

together with the corresponding sampling time. These included the sampling time, latitude, longitude, and monthly precipitation in the region. The data were processed and checked for quality. After this, 100 valid samples were obtained for modeling analysis.

Each sample therefore carried fourteen candidate inputs: the eight water-quality parameters listed above (WT, pH, COD_{Mn}, BOD_5, TP, TN, F^-, As), latitude and longitude, monthly precipitation, the sampling time, and the gross alpha and gross beta activities measured at the previous sampling time. The target is the current gross alpha activity. These fourteen inputs are thus a mix of water-quality, spatial, meteorological and temporal variables rather than water-quality indicators alone.

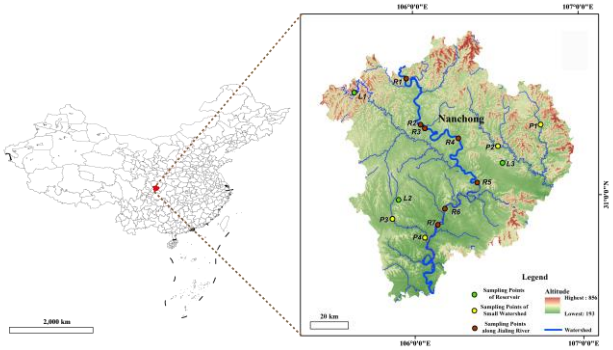


Figure 2. The distribution of water gross alpha radioactivity monitoring sites in Nanchong.

Table 1. The detailed information on environmental factors.

Environmental Factors	Abbreviations	Units	Methods (National Standard for Water Quality determination)
Gross alpha radioactivity	α	Bq/L	HJ 898-2017
Gross beta radioactivity	β	Bq/L	HJ 899-2017
Water temperature	WT	℃	GB 13195-91
Potential of hydrogen	pH	-	HJ 1147-2020
Permanganate index	COD_{Mn}	mg/L	GB 11892-89
Biochemical oxygen demand	BOD_5	mg/L	HJ 505-2009
Total phosphorus	TP	mg/L	GB 11893-89
Total nitrogen	TN	mg/L	HJ636-2012
Fluorine reagents	F^-	mg/L	HJ 84-2016
Arsenic	As	mg/L	HJ 694-2014

2.2. *Feature Dimension Reduction*

2.2.1. RF algorithm calculated feature importance

Features may have multicollinearity and information redundancy. To deal with this, dimensionality reduction was done before modeling. The RF algorithm was used for data filtering. The Gini coefficient was used to evaluate feature importance. Model accuracy was also used to find the optimal feature subset. The calculation steps are shown below (Jianjun Jiang, 2024; Karim, M. R, 2019; Zhibin Huang, 2022).

The Gini index was calculated as follows:

$$FI_s=\sum_{v=1}^{t}\sum_{\hat{v}\neq v}\lambda_{sv}\lambda_{s\hat{v}}=1-\sum_{v=1}^{t}\lambda_{sv}^2 \quad (1)$$

where t denotes the number of features, and λ_{sv} represents the proportion of the v-th feature at node s. The importance of feature F_i at node m is calculated as follows:

$$FIM_{is}^{(Gini)}=FI_s-GI_1-GI_2 \quad (2)$$

where GI_1 and GI_2 denote the Gini indices of the two new nodes generated after splitting. The importance of feature F_i in the k-th tree is then given by:

$$FIM_{ki}^{(Gini)}=\sum_{s\in M}FIM_{is}^{(Gini)} \quad (3)$$

where M represents the number of times feature F_i appears in the K-th tree. The importance of feature F_i across all k trees is then:

$$FIM_i^{(Gini)}=\sum_{k=1}^{K}FIM_{ki}^{(Gini)} \quad (4)$$

Finally, the weight of the i-th feature is normalized as follows:

$$FIM_i=\frac{FIM_i^{(Gini)}}{\sum_{i=1}^{m}FIM_i^{(Gini)}} \quad (5)$$

where m represents the total number of features. Based on Equations (1) through (5), the RF algorithm outputs the relative weight of each feature in the overall prediction task, reflecting the importance of the input features. The features were then sorted in descending order of importance. Subsequently, a series of feature subsets T_i, was constructed by incrementally adding features according to their rank.

2.2.2. Selection of the Optimal Feature Subset

Subsequently, each feature subset was independently input into a standard CNN prediction framework. The root mean square error (RMSE) of the model output was used to determine the optimal feature combination (Karim, M. R, 2019). A lower RMSE value indicates higher predictive accuracy, suggesting a better alignment between the feature subset and the model. This quantitative evaluation approach eliminates the subjectivity associated with

empirical selection, thereby improving the overall accuracy of the model.

2.3. *Establishment of a Model for Predicting Gross Alpha Radioactivity*

2.3.1. CNN Algorithm

A CNN is a representative feedforward deep neural network well-suited for processing grid-structured data. It has been widely applied in regression prediction tasks across various fields. Through hierarchical convolution operations, the model can automatically abstract and extract key features from raw inputs, thereby enabling efficient data-driven modeling.

The core component of a CNN is the convolutional layer. This layer slides a set of learnable local filters across the input data and performs discrete convolution operations to detect specific local patterns. This process can be expressed as follows:

$$O_{i,j,k}=\sum_{m=0}^{K_h-1}\sum_{n=0}^{K_w-1}\sum_{d=1}^{D}x_{i+m,j+n,d}\cdot w_{m,n,d,k}+b_k \quad (6)$$

Here, $O_{i,j,k}$ is referred to as the value at position (i,j) in the k-th output channel of the feature map, $x_{i+m,j+n,d}$ is considered to be the value at position (i+m, j+n) in the d-th input channel. The weight of the convolution kernel is given by $w_{m,n,d,k}$, corresponding to kernel position (m,n), input channel d, and output channel k. The parameters K_h and K_w denote the height and width of the filter, while D is the number of input channels. The bias term for output channel k is represented by b_k.

During forward propagation, the input data first passes through convolutional layers to detect local features. A ReLU activation function is then applied to introduce nonlinear transformations. Subsequently, pooling layers perform spatial downsampling to enhance feature invariance. Finally, fully connected layers integrate and compress the hierarchical features to produce the regression output.

Here the nine selected predictors form a one-dimensional vector, so each sample was arranged as a 9x1x1 tensor and passed through one-dimensional convolutions (Eq. 6). For the 9x1x1 input used here, Eq. (6) reduces to its one-dimensional special case (Kw = D = 1). The network stacks an input layer (9x1x1), a convolutional layer of sixteen 4x1 filters (same padding) with batch normalization and ReLU, a 3x1 max-pooling layer (stride 1), a second convolutional layer of thirty-two 4x1 filters with batch normalization and ReLU, a dropout layer (rate 0.16-0.18), a one-unit fully connected layer, and a mean-squared-error regression layer. Inputs and targets were scaled to [0,1] using training-fold statistics only, and predictions were mapped back before evaluation. We chose a 1-D CNN over a plain multilayer perceptron so that the shared convolution kernels could pick up local interactions between adjacent ranked features; even so, as the comparison below shows, its accuracy on this small tabular dataset is on par with the tree-ensemble regressors.

2.3.2. GWO optimized CNN

The GWO algorithm, proposed by Mirjalili *et al.* in 2014, is an efficient metaheuristic search strategy (F. Liu *et al.*, 2024). Compared with other optimization methods, GWO features a simple structure, requires minimal parameter settings, and possesses strong global exploration capabilities. These characteristics make it particularly suitable for solving complex non-convex optimization problems, such as hyperparameter tuning in deep neural networks. Within the GWO algorithm, population individuals are stratified into four tiers based on fitness: the fittest individual is designated as the alpha wolf (α), the second-fittest as the beta wolf (β), the third-fittest as the delta wolf (δ), and all remaining individuals as omega wolves (ω). The hunting process primarily consists of two core stages: encirclement and attack. The encirclement behavior plays a crucial role in guiding the population toward convergence in the optimal region (Bwapwa & Mukuna, 2026; Kaissar *et al.*, 2025). The corresponding mathematical expressions are presented in Equations (7–11).

$$D=\left|\vec{C}\cdot\vec{X}_p t-\vec{X}t\right| \quad (7)$$

$$\vec{X}(t+1)=\vec{X}_p t-\vec{A}\cdot\vec{D} \quad (8)$$

$$\vec{A}=2\vec{a}\cdot\vec{r}_1-\vec{a} \quad (9)$$

$$\vec{C}=2\cdot\vec{r}_2 \quad (10)$$

$$\vec{a}=2-t\cdot\left(\frac{2}{T_{max}}\right) \quad (11)$$

In the GWO algorithm, t denotes the current iteration count, $\vec{X}_p$ represents the position of the prey, and $\vec{X}$ denotes the current position of a wolf. The convergence factor $\vec{a}$ decreases linearly from 2 to 0 over the course of the iterations, thereby balancing global exploration with local exploitation. The parameters $\vec{r}_1$ and $\vec{r}_2$ are random vectors distributed within the interval [0,1]. Equation (7) describes the distance between the prey and the wolf after the t-th iteration, while Equation (8) specifies the updated position of the wolf after the (t+1)-th iteration. Together, these equations describe the process by which the wolf pursues its prey and adjusts its position accordingly. Once the prey is locked onto, the wolf rapidly approaches and initiates an attack. The hunting process is mathematically expressed in Equations (12–14).

$$\vec{D}_\alpha=\left|\vec{C}_1\cdot\vec{X}_\alpha-\vec{X}\right|,\vec{D}_\beta=\left|\vec{C}_2\cdot\vec{X}_\beta-\vec{X}\right|,\vec{D}_\delta=\left|\vec{C}_3\cdot\vec{X}_\delta-\vec{X}\right| \quad (12)$$

$$\overrightarrow{X_1}=\left(\vec{X}_\alpha-\vec{A}_1\cdot\vec{D}_\alpha\right),\overrightarrow{X_2}=\left(\vec{X}_\beta-\vec{A}_2\cdot\vec{D}_\beta\right),\overrightarrow{X_3}=\left(\vec{X}_\delta-\vec{A}_3\cdot\vec{D}_\delta\right) \quad (13)$$

$$\vec{X}(t+1)=\frac{\overrightarrow{X_1}+\overrightarrow{X_2}+\overrightarrow{X_3}}{3} \quad (14)$$

In the algorithm, the intervals between an omega wolf (ω) and the three leading wolves (α, β, δ) are quantified by the distance vector $\vec{D}_\alpha,\vec{D}_\beta,\text{and}\,\vec{D}_\delta$. Three candidate positions, denoted as X_1, X_2 and X_3, are then calculated according to the guidance from each leading wolf. The final position of the omega wolf is computed as the average of X_1, X_2 and X_3.

At each iteration, the three wolves with the highest fitness values are dynamically chosen as the new α, β and δ wolves. Based on their positions, the probable location of the prey is estimated. All ω wolves then update their positions around this estimated location. As a result, the entire pack is driven toward convergence in the optimal solution region.

In this study, the GWO algorithm is used to search for optimal CNN hyperparameters globally and automatically. This helps avoid the problems that come with manual tuning. With this method, key hyperparameters that directly affect the training dynamics and generalization of the CNN, such as the learning rate, the number of training iterations, and the regularization parameters, are optimized. As a result, prediction accuracy is improved. At the same time, computational time is greatly reduced.

The GWO search used five wolves over thirty iterations and tuned three hyperparameters within preset ranges: learning rate [1e-3, 1e-2], training epochs [500, 800], and L2 coefficient [1e-4, 1e-3]. Each candidate was scored by its validation RMSE on an inner hold-out set. The search settled within the thirty-iteration budget (**Figure 8**) on a learning rate of 1.0e-3, 707 epochs and an L2 coefficient of 4.5e-4, which were then used to train the final CNN with the Adam optimizer.

2.3.3. Evaluation Indicators

Two error metrics and their ratio are used to evaluate the performance of the predictive model(Uzbas & Kocaman, 2025). These are the RMSE, the mean absolute error (MAE) and the RMSE-to-MAE ratio. The calculation procedures for RMSE and MAE are shown in Equations (15 and 16).

$$MAE=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \quad (15)$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2} \quad (16)$$

Where n denotes the sample size, yi represents the observed measurement, and $\hat{y}_i$ indicates the model's predicted value.

With only 100 samples, we assessed the models by repeated cross-validation rather than a single split: five-fold cross-validation was run ten times (50 folds in total), and we report the mean and standard deviation of MAE and RMSE across folds. Normalization was fitted on the training folds alone to avoid information leakage. To test whether the differences were meaningful, a paired Wilcoxon signed-rank test was applied to the per-fold RMSE of the proposed model against each baseline, with $p < 0.05$ taken as significant.

3. Results and Discussion

3.1. *Dimensionality Reduction*

3.1.1. RF Importance Ranking

To enhance the prediction accuracy of the model, in this study, when constructing the input features, the gross alpha radioactivity and gross beta radioactivity measurement values of the previous time point at each monitoring point were included for the time series-related variables. All the features were input into the RF algorithm, and the output results were quantitatively sorted from largest to smallest to obtain the ordered feature set T_i (Karim, M. R, 2019). The sorting result is shown in **Figure 3**.

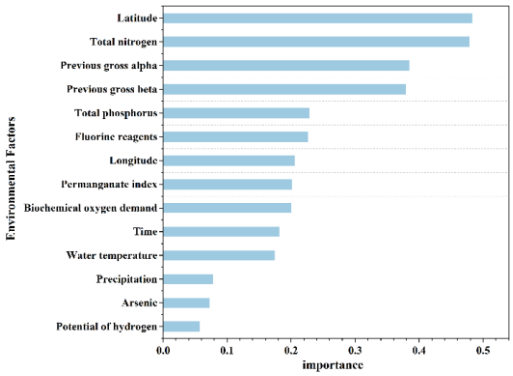


Figure 3. Importance of water quality indicators used as input features for predicting gross alpha radioactivity in Nanchong City (study period).

Table 2. Comprehensive Evaluation Results for Optimal Feature Subsets.

D_w	RMSE					average
	$test_1$	$test_2$	$test_3$	$test_4$	$test_5$	
D_1	0.0205	0.0200	0.0193	0.0192	0.0202	0.01983
D_2	0.0169	0.0185	0.0176	0.0178	0.0160	0.01738
D_3	0.0166	0.0174	0.0172	0.0163	0.0159	0.01668
D_4	0.0154	0.0162	0.0147	0.0158	0.0152	0.01546
D_5	0.0142	0.0154	0.0156	0.0150	0.0141	0.01484
D_6	0.0197	0.0193	0.0189	0.0201	0.0191	0.01943
D_7	0.0162	0.0175	0.0178	0.0163	0.0153	0.01663
D_8	0.0195	0.0188	0.0205	0.0197	0.0188	0.01945
D_9	**0.0131**	**0.0144**	**0.0137**	**0.0139**	**0.0142**	**0.01386**
D_{10}	0.0215	0.0208	0.0198	0.0184	0.0193	0.01998
D_{11}	0.0164	0.0168	0.0179	0.0173	0.0166	0.01699
D_{12}	0.0197	0.0203	0.0217	0.0196	0.0203	0.02033
D_{13}	0.0178	0.0164	0.0186	0.0176	0.0188	0.01783
D_{14}	0.0169	0.0177	0.0165	0.0154	0.0154	0.01637

3.1.2. Selection of the Best Feature subset

For each candidate feature set containing the top n features from T_i, a standard CNN model was trained and evaluated. This procedure was carried out five times per feature set. The five resulting RMSE values were then averaged, and the subset with the lowest average RMSE was selected as the optimal feature set. The ordered feature set is denoted as $D_w = \{T_i \mid i = 1, 2, \ldots, w\}$, where w is the number of selected features. The corresponding results are presented in **Table 2**.

When measurements are taken along the north–south line of the Jialing River main stem, higher values are seen in the upstream section than in the downstream section. This is interpreted as evidence that downstream waters have a certain capacity to purify upstream waters. Other factors may also play important roles, including the geological background of the watershed, the spatial distribution of rock types, and the migration and accumulation of radionuclides (Ali, Abbady, *et al.*, 2025).

The geochemical interpretations that follow are hypotheses suggested by the data-driven RF ranking rather than conclusions from direct mechanistic measurement, and confirming them would require targeted hydrogeochemical study.Among the nutrient indicators, TN and TP were ranked high in importance. From a geochemical perspective, salts introduced by human activities may influence gross alpha radioactivity concentrations. This can happen through changes in the redox potential of the water, or through competitive adsorption and complexation with alpha-emitting nuclides such as uranium (U) and radium (Ra).

F^- was identified as a highly important feature. It may form complexes with alpha-emitting nuclides or become involved in coprecipitation. This suggests that ligand complexation can influence the migration and chemical form of gross alpha radioactivity in water.

In addition, both the previous gross beta and previous gross alpha measurements were found to have high feature importance. This observation supports the hypothesis that alpha and beta radioactivity in the region may originate from a common source.

It was found that certain routine water quality indicators exhibited higher feature importance than the radionuclides themselves. This finding suggests that routine indicators may give earlier and more sensitive signals than radionuclides. They can reflect the hydrogeochemical processes that cause radioactivity changes. Accordingly, nine indicators were chosen as the optimal feature subset for later modeling. namely latitude, longitude, COD_{Mn}, BOD_5, TP, TN, F^-, previous gross beta, and previous gross alpha. With this subset, the current gross alpha radioactivity can be predicted using existing, rapidly measurable routine water quality indicators, along with previous gross beta radioactivity and previous gross alpha radioactivity.

3.2. *Model Prediction*

3.2.1. Comparison of different models

To systematically evaluate the predictive capability of the RF-GWO-CNN model for low-level environmental radioactivity, this study compared BP neural networks, SVM, CNN, and the GWO-optimized CNN model. To ensure comparability among the models, the optimal feature subset extracted in Section 3.1 was first split into training and testing sets in an 8:2 ratio before prediction. The visualization results of the predictions are shown in Figure **4**. On this split the RF-GWO-CNN prediction tracked the observations reasonably well in trend, high-value fit and error spread; Section 3.2.2 quantifies how much of that advantage survives once resampling variability is taken into account.

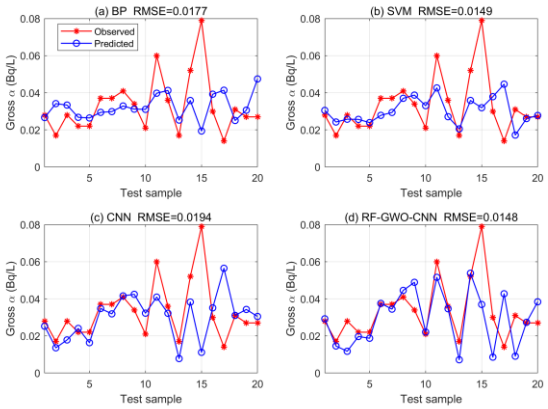


Figure 4. Predicted and observed gross alpha radioactivity using four models: (a) BP, (b) SVM, (c) CNN, and (d) RF-GWO-CNN.

On this split the four models reached broadly similar accuracy (panel RMSE values in **Figure 4**); in line with the cross-validation in Section 3.2.2, none was decisively better and the order of the leading models shifted from split to split, so any single-split ranking should be read with caution given the small sample. This may reflect the presence of complex nonlinear relationships among the variables. Furthermore, the limited sample size (100 samples) poses a challenge for training deep learning models and should be considered when interpreting the results.

For the BP neural network in **Figure 4(a)**, the predictions were close in value but did not follow the true trend. The consistency of the trend was low, and high-value samples were poorly estimated.

The SVM predictions in **Figure 4(b)** followed the trend more closely than BP. Still, the model's generalization ability did not improve clearly. The main reason is that the training data had only a few high-value samples. Therefore, the model did not learn enough from them. This led to local underfitting in the high-value regions.

As shown in **Figure 4(c)**, the CNN's local feature extraction made the predictions more physically reasonable and trend-consistent. However, the predicted values still oscillated around the true values. More work is needed to improve the prediction accuracy.

The GWO algorithm globally tunes the CNN regularization settings. As seen in **Figure 4(d)**, the RF-GWO-CNN model captured the overall trend but, like the other models, produced larger errors at a few extreme low and high

values; its single-split accuracy was similar to that of the other models, in line with the cross-validated results in Section 3.2.2.

3.2.2. Reliability Assessment

A single 8:2 hold-out split initially suggested a considerably lower error for RF-GWO-CNN (MAE 0.0081 Bq/L, RMSE 0.0095 Bq/L). Since n = 100 and a single split is very sensitive to sampling, we re-evaluated every model with ten repeats of five-fold cross-validation (50 folds; **Table 3, Figures 5 and 7**). The cross-validated gaps are far smaller and mostly overlap: RF-GWO-CNN averaged an RMSE of 0.0169 (SD 0.0035) Bq/L and MAE of 0.0126 (SD 0.0022) Bq/L, effectively level with the standard CNN (0.0172 / 0.0126) and the random forest, which had the lowest mean errors (0.0166 / 0.0122), and ahead of BP (0.0184 / 0.0135) and SVM (0.0188 / 0.0146). In the paired Wilcoxon test it beat only SVM significantly ($p < 0.001$); its differences from BP, RF, gradient boosting and the standard CNN were not significant ($p = 0.15$, 0.45, 0.19 and 0.68). For a dataset this small, the proposed model is thus competitive with the best baselines rather than clearly superior.

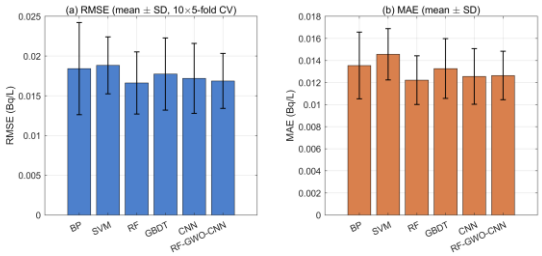


Figure 5. Cross-validated RMSE (a) and MAE (b) for the six models (mean and standard deviation over 50 folds; error bars denote SD).

Table 3. Cross-validated performance (mean and SD over 50 folds, 10x5-fold CV) with a paired Wilcoxon signed-rank test versus RF-GWO-CNN.

Model	MAE (Bq/L)	RMSE (Bq/L)	p vs RF-GWO-CNN
BP	0.0135 (0.0030)	0.0184 (0.0058)	0.15
SVM	0.0146 (0.0023)	0.0188 (0.0036)	<0.001
RF	0.0122 (0.0022)	0.0166 (0.0039)	0.45
Gradient boosting	0.0133 (0.0027)	0.0177 (0.0045)	0.19
CNN	0.0126 (0.0025)	0.0172 (0.0044)	0.68
RF-GWO-CNN	0.0126 (0.0022)	0.0169 (0.0035)	reference

MAE is the average absolute difference between predicted and observed values. It gives a general measure of prediction accuracy. However, MAE is not sensitive to occasional large errors (Li *et al.*, 2019). Therefore, it does not fully reflect the fluctuations in prediction results.

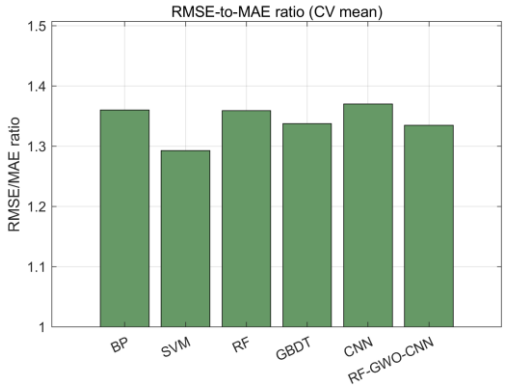


Figure 6. RMSE-to-MAE ratio for the six models (cross-validated means).

In contrast, RMSE is calculated by squaring the errors before averaging and taking the square root. This makes RMSE more sensitive to large errors. As a result, RMSE is better at measuring the spread of prediction errors and the uncertainty of the model (T. Liu *et al.*, 2025; Chang *et al.*, 2024).

The RMSE-to-MAE ratio (**Figure 6**) reflects the error distribution and generalization ability of a model. A high ratio means the errors are more widespread, often due to a few large outliers. This usually indicates unstable predictions and poor generalization. A low ratio means the errors are more concentrated, which suggests that the model is both accurate and stable (Karunasingha, 2022).

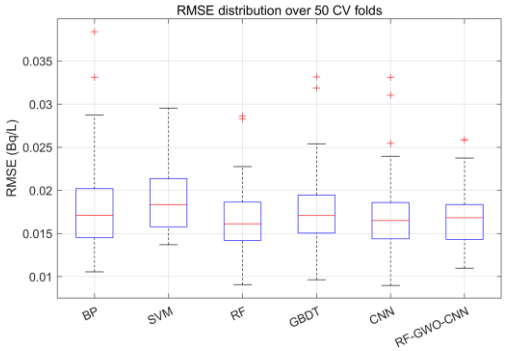


Figure 7. Distribution of RMSE over the 50 cross-validation folds for the six models; the boxes overlap substantially, in line with the non-significant pairwise differences.

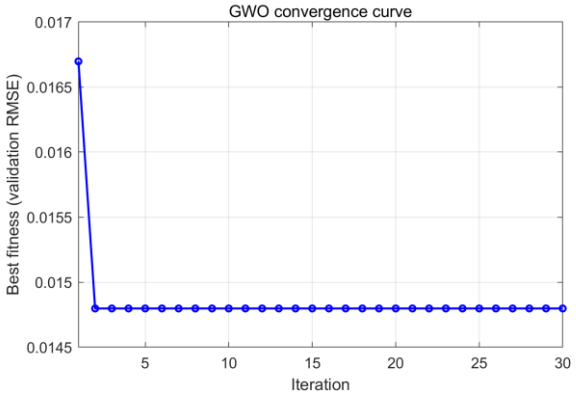


Figure 8. Convergence curve of the Gray Wolf Optimization search for the CNN hyperparameters.

On the single hold-out split the RF-GWO-CNN model gave a low RMSE-to-MAE ratio (about 1.2), indicating a concentrated error distribution; under cross-validation the leading models (RF, CNN and RF-GWO-CNN) show similar ratios (**Figure 7**), consistent with their comparable overall accuracy.

3.2.3. Optimal Model Selection

Taken together, Sections 3.2.1 and 3.2.2 place RF-GWO-CNN among the best models, with accuracy close to the random forest and standard CNN and a significant edge over SVM.It performs well in the low-concentration range and has a more concentrated error distribution. This indicates higher stability and reliability. Therefore, the model is suitable for early warning of low-level gross alpha radioactivity in water.

3.3. *Innovation, Limitations, and Future Work*

Most of the selected routine water quality indicators can be measured quickly or online. This time advantage can be used to build an early warning system for radioactivity anomalies. The system would take historical radioactivity data, current water quality indicators, and spatial coordinates as inputs. When an anomaly is detected in these indicators, the system can quickly predict a possible radioactivity event. This allows radiation incidents to be detected much earlier than with traditional emergency methods, leaving more time for response. In this way, passive waiting is replaced by active early warning. This may help make better use of limited monitoring resources and improve the speed and efficiency of radiation risk warning.

However, some uncertainties remain. These include limited sample variety, gaps in temporal sampling, statistical fluctuations in radionuclides, and manual analysis errors. Therefore, the model's applicability needs further testing. The BOD_5 indicator used in this study is not ideal in terms of timeliness, but it was still included because of its importance.

The small dataset brings further caveats. With only 100 samples, overfitting is a real risk for a deep model; we limited it by cutting the inputs to nine RF-selected features, applying dropout and L2 regularization, and reporting repeated cross-validation rather than a single split. The cross-validated results still show tree-ensemble regressors matching the deep model, which is the usual outcome for small tabular problems. Because lagged gross alpha and gross beta values are used as inputs, we were careful about temporal leakage: only past measurements enter the feature vector, the target is never fed back as an input, and normalization is fitted on the training folds only. All data also come from a single basin (the Jialing River at Nanchong), so the model's transferability to other regions or hydrogeochemical settings remains untested and should be regarded as a limitation until it is checked on independent basins.

Future work will focus on two areas. First, we will develop continuous monitoring models that are closer to real accident conditions. Second, we will look for input features with stronger links to radioactivity. These efforts will help confirm the practical value of the proposed early warning system.

4. Conclusion

In environmental radioactivity early warning, gross alpha radioactivity in water may correlate with routine water quality indicators. However, when data is scarce and statistical fluctuations are large, traditional models cannot accurately predict. To solve this, an RF-GWO-CNN model based on RF feature dimensionality reduction is proposed in this study. We analyzed 100 datasets and obtained the following findings:

1) Using RF feature importance evaluation, routine water quality indicators closely associated with gross alpha radioactivity can be identified, while redundant features can be eliminated. This approach not only enhances the generalization capability of the model but also provides a statistical basis for identifying the key influencing factors of gross alpha radioactivity.
2) A nonlinear relationship exists between routine water quality indicators and gross alpha radioactivity in water, which may be related to hydrogeochemical processes. These indicators can serve as key predictors for estimating gross alpha radioactivity at the environmental level.
3) For gross alpha radioactivity, RF-GWO-CNN gave a cross-validated MAE of about 0.013 Bq/L and RMSE of about 0.017 Bq/L, close to a random-forest regressor and a standard CNN and significantly better than SVM. Its practical appeal is that the RF feature selection and GWO tuning make the pipeline interpretable and easy to redeploy, giving Nanchong a workable basis for early warning of gross alpha radioactivity, even if its accuracy edge over simpler tree ensembles is modest on the present small dataset.

In conclusion, the RF-GWO-CNN model built in this study offers a practical way to predict gross alpha radioactivity accurately in data-scarce river basins. It can also serve as a reference for setting up regional water risk management and early warning systems under radiation accident conditions.

Acknowledgments

The authors thank the Nanchong Ecological and Environmental Monitoring Central Station of Sichuan Province for providing access to the data used in this study. This research was funded by the Science and Technology Plan Project of Nanchong City, grant number 23YYJCYJ0055. The views and conclusions presented herein are solely those of the authors and do not necessarily reflect the official position of the station. The authors also appreciate the valuable discussions and suggestions contributed by their colleagues.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

A. Helaly, M., Rady, S., Mabrouk, M., M. Aref, M., Villarroya, S., Cotos, J. M., & Mera, D. (2025). Advancements in water quality prediction: A practical review of machine learning and

deep learning approaches. Cluster Computing, 28(9), 598–614. https://doi.org/10.1007/s10586-025-05221-3

Abobakr Yahya, A. S., Ahmed, A. N., Binti Othman, F., Ibrahim, R. K., Afan, H. A., El-Shafie, A., Fai, C. M., Hossain, M. S., Ehteram, M., & Elshafie, A. (2019). Water Quality Prediction Model Based Support Vector Machine Model for Ungauged River Catchment under Dual Scenarios. *Water*, *11*(6), 1231. https://doi.org/10.3390/w11061231

Al-Amir, S. M., Al-Hamarneh, I. F., Al-Abed, T., & Awadallah, M. (2012). Natural radioactivity in tap water and associated age-dependent dose and lifetime risk assessment in Amman, jordan. *Applied Radiation and Isotopes: Including Data, Instrumentation and Methods for Use in Agriculture, Industry and Medicine*, *70*(4), 692–698. https://doi.org/10.1016/j.apradiso.2011.12.002

Ali, K., Abbady, A. E.-B., Abu-Taleb, A., & Harb, S. (2025). Assessment of radioactive substance transfer and its ecological and health impacts on the nasser lake ecosystem. *Scientific Reports*, *15*(1), 26115. https://doi.org/10.1038/s41598-025-09342-y

Ali, K., Matar, Z. S., Walther, C., Salah El-Din, K., Harb, S., Kilany, M., & Moubark, K. (2025). Radionuclide distribution and hydrochemical controls in groundwater of the nile valley, upper egypt: Health and environmental implications. *Water*, *17*(18), 2730. https://doi.org/10.3390/w17182730

Bwapwa, J. K., & Mukuna, J. G. (2026). Metaheuristic optimization of treated sewage wastewater quality parameters with natural coagulants. *Water*, *18*(8), 885–918. https://doi.org/10.3390/w18080885

Chang, C.-Y., Peng, C.-H., Chen, F.-Y., Huang, L.-Y., Kuo, C.-H., Chu, T.-W., & Liang, Y.-J. (2024). The risk factors determined by four machine learning methods for the change of difference of bone mineral density in post-menopausal women after three years follow-up. *Scientific Reports*, *14*(1), 23234. https://doi.org/10.1038/s41598-024-73799-6

Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. *Applied Sciences*, *10*(17), 5776. https://doi.org/10.3390/app10175776

Ghisellini, P., Passaro, R., & Ulgiati, S. (2025). Assessing the environmental sustainability and justice dimensions of nuclear electricity under circular economy and energy transition frameworks. Journal of Cleaner Production, 491, 144818. https://doi.org/10.1016/j.jclepro.2025.144818

He, H., Qian, Y., Yuan, X., Hu, J., Zhang, H., & Xia, J. (2023). Temporal variation trend and influencing factors of water quality in the nanchong section of the jialing river mainstream. *Desalination and Water Treatment*, *286*, 282–292. https://doi.org/10.5004/dwt.2023.29336

Hou, H., Yin, J., Qiu, X., Wu, J., Yang, Y., Tang, D. W. S., Wang, Y., Wei, Z., & Xu, C. (2025). Water heavy metal prediction using a CBM model based on hybrid metaheuristic algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, *63*, 1–31. https://doi.org/10.1109/TGRS.2025.3602639

Huang, Z. (2022). Modern decision tree models and programming practices: From traditional decision trees to deep decision trees. China Machine Press.

IAEA. (2022, January 28). *Fukushima daiichi ALPS treated water discharge* [Web page]. http://www.iaea.org/topics/response/fukushima-daiichi-nuclear-accident/fukushima-daiichi-alps-treated-water-discharge

Jiang, J. (2024). Big Data Analytics Technology. Xidian University Press.

Kaissar, A., Nassif, A. B., Soudan, B., & Injadat, M. (2025). Enhancing CNN-based network intrusion detection through hyperparameter optimization. *Intelligent Systems with Applications*, *26*, 200528. https://doi.org/10.1016/j.iswa.2025.200528

Karim, M. R. (2019). *Scala machine learning projects: Build real-world machine learning and deep learning projects with Scala* (Shiyong Zhang & Bo Liu, Trans.). China Machine Press.

Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, *585*, 609–629. https://doi.org/10.1016/j.ins.2021.11.036

Klingelhöfer, D., Braun, M., Oremek, G. M., Brüggmann, D., & Groneberg, D. A. (2024). Global research on nuclear energy in the context of health and environmental risks, considering economic interests. *WIREs Energy and Environment*, *13*(1), e497. https://doi.org/10.1002/wene.497

Ladi, S. K., Panda, G. K., Dash, R., Ladi, P. K., & Dhupar, R. (2022). A novel Grey Wolf optimisation based CNN classifier for hyperspectral image classification. *Multimedia Tools and Applications*, *81*(20), 28207–28230. https://doi.org/10.1007/s11042-022-12628-2

Li, B., Liu, Y., Wang, X., Fu, Q., & Lv, X. (2019). Application of the orthogonal polynomial fitting method in estimating PM2.5 concentrations in central and southern regions of China. *International Journal of Environmental Research and Public Health*, *16*(8), 1418. https://doi.org/10.3390/ijerph16081418

Liu, F., Ding, N., Zheng, G., & Xu, J. (2024). Carbon emission prediction and reduction analysis of wastewater treatment plants based on hybrid machine learning models. *Environmental Engineering Research*, *30*(2), 240403–0. https://doi.org/10.4491/eer.2024.403

Liu, T., Jiao, W., Huang, Z., Yu, J., Zhang, X., & Li, L. (2025). A short-term load forecasting framework for air conditioning system based on model stacking. *Scientific Reports*, *15*(1), 36892. https://doi.org/10.1038/s41598-025-20929-3

Luo, X., Guo, Q., Tian, Y., Cao, J., & Luo, G. (2025). Remote sensing inversion model of water quality parameters based on BP neural network and spatial distribution analysis in the middle reaches of the Yangtze River basin in China. *Journal of the Indian Society of Remote Sensing*, *53*(8), 2535–2557. https://doi.org/10.1007/s12524-025-02149-4

Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling and Software*, *15*(1), 101–124. https://doi.org/10.1016/S1364-8152(99)00007-9

Muñoz-Alegría, J. A., Núñez, J., Oyarzún, R., Chávez, C. A., Arumí, J. L., & Rodríguez-López, L. (2025). A bibliometric-systematic literature review (B-SLR) of machine learning-based water quality prediction: Trends, gaps, and future directions. *Water*, *17*(20), 2994. https://doi.org/10.3390/w17202994

Sathya, P., Subramanian, P., Soniya, M., Pugalenthi, R., & M, S. P. V. (2024). Water quality index prediction and classification using hyperparameter tuned deep learning approach. Global NEST Journal, 26(5), 1-8. https://doi.org/10.30955/gnj.005821

Tăban, C. I., Benedek, A. M., Stoia, M., Cocîrlea, M. D., & Oancea, S. (2023). A Multivariate Model of Drinking Water Quality Based on Regular Monitoring of Radioactivity and Chemical

Composition. *Applied Sciences*, *13*(18), 10544. https://doi.org/10.3390/app131810544

Uzbas, B., & Kocaman, S. (2025). A machine learning-based modeling approach for dye removal using modified natural adsorbents. *Journal of Chemical Information and Modeling*, *65*(16), 8486–8496. https://doi.org/10.1021/acs.jcim.5c01016

Visu, P., Rajesh Khanna, M., Sundara Rajulu, N., & Subramanian, P. (2024). Air pollution prediction using attention module with CNN-OptBiLSTM. Global NEST Journal, 26(6), 1-9. https://doi.org/10.30955/gnj.05823

Wang, X., Shi, F., Yu, L., & Li, Y. (2013). MATLAB neural network: 43 case analyses. Beihang University Press.

Yang, S., Zhong, S., & Chen, K. (2024). W-WaveNet: A multi-site water quality prediction model incorporating adaptive graph convolution and CNN-LSTM. *PLoS One*, *19*(3), e0276155. https://doi.org/10.1371/journal.pone.0276155

Yang, W., Xie, S., Zhang, Z., Hu, J., Zhang, L., Lei, X., Zhong, L., Hao, Y., & Shi, F. (2019). Characteristics and sources of carbonaceous aerosol across urban and rural sites in a rapidly urbanized but low-level industrialized city in the sichuan basin, china. *Environmental Science and Pollution Research*, *26*(26), 26646–26663. https://doi.org/10.1007/s11356-019-05242-7

Zhang, D. (2021). MATLAB R2020a neural network typical case analysis. Publishing House of Electronics Industry.

Zhu, J.-J., Yang, M., & Ren, Z. J. (2023). Machine learning in environmental research: Common pitfalls and best practices. *Environmental Science & Technology*, *57*(46), 17671–17689. https://doi.org/10.1021/acs.est.3c00026

Zhuo, Y. (2018). Analysis on the current situation and countermeasures of ecological barrier construction in the middle reaches of Jialing River in Nanchong City. China Green Pictorial, (7), 124.