# Hybrid Machine Learning Approaches for Accurate Forecasting of Total Dissolved Solids: Case Study of a Chinese River

***Authors Names***

*Balsam Mustafa Shafeeq[1], Lamiaa Abdul-Jabbar Dawod[2] Mustafa Abdul Jabbar Dawood[3],Waleed Ahmed Hassen Al-Nuaami[4]*

Publication data: 25 / 12 /2025

***Keywords:*** *Algorithm, Intelligent model, Meta engineering, Prediction, water quality*

**ABSTRACT**

Surface water management is one of the important factors of water quality in industrial, agricultural and urban basins. One of the most important quality indictors is Total Dissolved Solids (TDS). Also, sampling for TDS is expensive, so metaheuristic methods are very suitable and cost-effective. In this study, SVM-IWO and SVM-TLBO metaheuristic models were used to simulate TDS in the Kim-Tin River in China. The monthly measured temperature, pH , Salinity turbidity and Total Dissolved Solids (TDS) from 2000-2023 data were used . The evaluation criteria of the coefficient of determination and root mean square error were used to compare the results. The results showed that the SVM-IWO metaheuristic method provided a better simulation in the accuracy section than the SVM-TLBO method ($R^2$ = 0.74 RMSE=63 mg/l). In general, there is a little difference between the Total Dissolved Solids (TDS) simulation of these two metaheuristic models. Either model can also be used to simulate TDS in the river.

## 1. Introduction

Considering the vital role of water quality in managing river ecosystems, evaluating total dissolved solids (TDS) remains a crucial task. Rivers, as major sources for domestic and industrial water use, require continuous quality monitoring, particularly under conditions of drought and ongoing urban and rural development. In this context, several recent studies have sought to enhance TDS prediction accuracy using intelligent and hybrid modeling approaches. For example, Akhoni Pourhosseini et al. (2023) examined the Babolrud River in Iran and applied several SVM-based hybrid models—SVM-CA, SVM-HS, and SVM-TLBO—using monthly measurements of pH, Ca, Mg, $HCO_3$, Na, $SO_4$, Cl, and TDS. Their findings indicated that the SVM-TLBO model produced the most reliable predictions. Similarly, Sayadi et al. (2024) employed the Grey Wolf Optimization algorithm combined with a Kernel Extreme Learning Machine (GWO-KELM) to model TDS, achieving an $R^2$ of 0.974 and RMSE of 60.13, outperforming SVM and ANN models. In another study, Jamshidzadeh et al. (2024) explored TDS and electrical conductivity (EC) in coastal aquifers using CNNE, LOST, and GPRE hybrid models. Their results revealed that the integrated CNNE-LSTM-GPR approach captured complex nonlinear relationships among water quality parameters and yielded more accurate predictions.

Memar et al. (2025) combined SVM with the COA algorithm to estimate pollution levels in Iran's Jajrud River using hadrochemical data from multiple stations, reporting that SVM-COA outperformed LSSVM, especially at Sharifa bad. Likewise, Singh (2025) applied several machine learning models, including Decision Tree, Logistic Regression, and SVM, to classify water quality in Telangana, India, finding that hybrid approaches enhanced model accuracy. Mathaba (2023) evaluated SVM applications for water treatment and resource management and emphasized the need for better data reliability and practical validation of intelligent methods. Earlier, Khatibi et al. (2017) modeled Bear River flow in the United States with ANN-based hybrid structures (MLP-LM and MLP-FFA), showing that MLP-LM produced superior results. Furthermore, Barzegar et al. (2023) assessed groundwater quality in Iran's Yazd Province using ANFIS, SVM, and ANN trained with algorithms such as PSO, GWO, CSO, and SA. Their findings, consistent with Ghorbani and Pamucar (2025), showed that hybrid ANFIS models—especially ANFIS-MFO and ANFIS-CSO—offered the highest predictive accuracy.

[1]Technical College of Management –Baghdad, Middle Technical University, Iraq, balsammustafa95@mtu.edu.iq

[2]Middle technical university /College of Health and medical Techniques/Baghdad Department of Optics Techniques, lamya.abd@mtu.edu.iq

[3] University of Baghdad / college of administration and economics/Statistics Department, mostafa.abduljabar1201a@coadec.uobaghdad.edu.iq

[4] Department of Biology, College of Education for Pure Sciences, University of Diyala., purecomp.waleed.hassan@uodiyala.edu.iq

Finally, Gulati et al. (2025) investigated groundwater pollution in West Bengal, India, using SVMs, ANNs, and ANFIS, and concluded that ANN achieved the lowest error values (Yildirim et al., 2025).

The main objective of this study is to determine the water quality of the Kim Tin River in Hong Kong, one of the major rivers in the basin. To determine TDS in this study, innovative and meta-heuristic methods SVM-IWO and SVM-TLBO were used, utilizing monthly data on temperature, pH, salinity, and turbidity for the years 2000-2023.

## 2. Materials and Methods

### 2.1 Study area and data required

The case study is Kam Tin River, it is located on the southeast coast of China. it is in the northwest of the Hong Kong. Hong Kong location is on the east bank of pearl river, it is the best position for Kim Tin River. (Goggins et al., 2012). The river basin is equal to 44.3 square kilometers and its length is 13 kilometers. (Liu et al., 2023). It is an important hub in east Asia, because of major cities such as Guangzhou and Shanghai, as well as nearby Taiwan. The Kam-Tin River is one of the important rivers of Hong Kong for water resource management in terms of irrigation and drinking water supply. And Kam-Tin River is located 22.3193° N latitude, 114.1694° E longitude coordinated. See figure1.

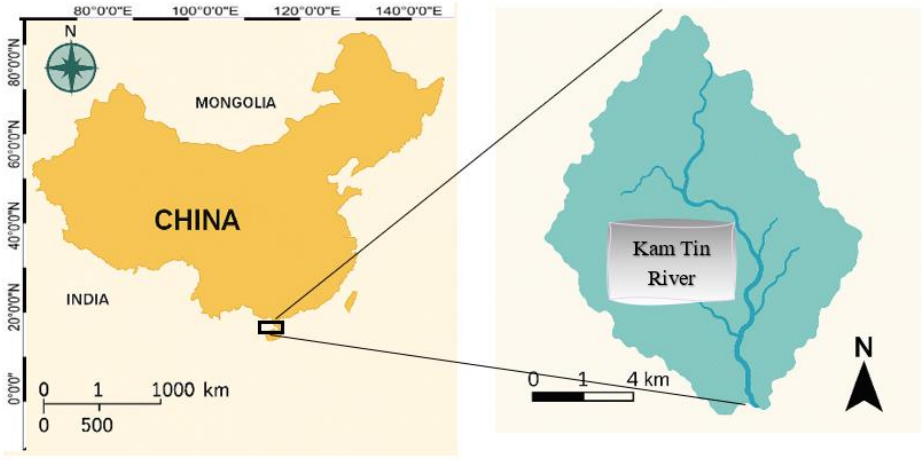


**Figure 1 illustrates the geographical location of the Kim Tin River.**

Data used in study were obtain monthly data including turbidity, temperature, pH, and salinity and TDS from 2000 -2023. Data were gathered from the Hong Kong database (https://data.gov.hk/). After deleted outlifted and noisy data, for estimate TDS, used input data including turbidity, temperature, pH, and salinity. Metaheuristic SVM- IWO and SVM-TLBO were used for simulation. For monthly TDS prediction, 70% of the data was allocated to model training and 30% to testing. all modeling and prediction processes were used in MATLAB. Table1.show statistical of water quality parameters in the Kim-Tin River.

. **Table1. Statistics of Water Quality Parameters in the Kim Tin River**

| | Kam Tin | | | |
|---|---|---|---|---|
| Statistical | Max | Mean | Min | STD |
| TDS (mg/L) | 570 | 246.1 | 110 | 89.78 |
| pH | 9.9 | 7.3 | 4.7 | 0.36 |
| Salinity (psu) | 0.4 | 0.16 | 0 | 0.068 |
| Turb (NTU) | 195.8 | 21.6 | 2.4 | 29.09 |
| Temperature ($^oC$) | 33.3 | 24.78 | 13.9 | 4.19 |

T=Teacher$\mu$new = T (1)

$$x_{inew} = x_i + r(\mu_{new} - Tf\mu)$$

TF=Teaching Factor$\in \{1,2\}$

Learning Phase

In this phase, knowledge is transferred through interaction between students. Figure2. SVM-TLBO model structure.

$$x_{inew} = x_i + r(x_j - x_i) \quad (2)$$

$$x_{inew} = x_i + r(x_i - x_j) \quad (3)$$

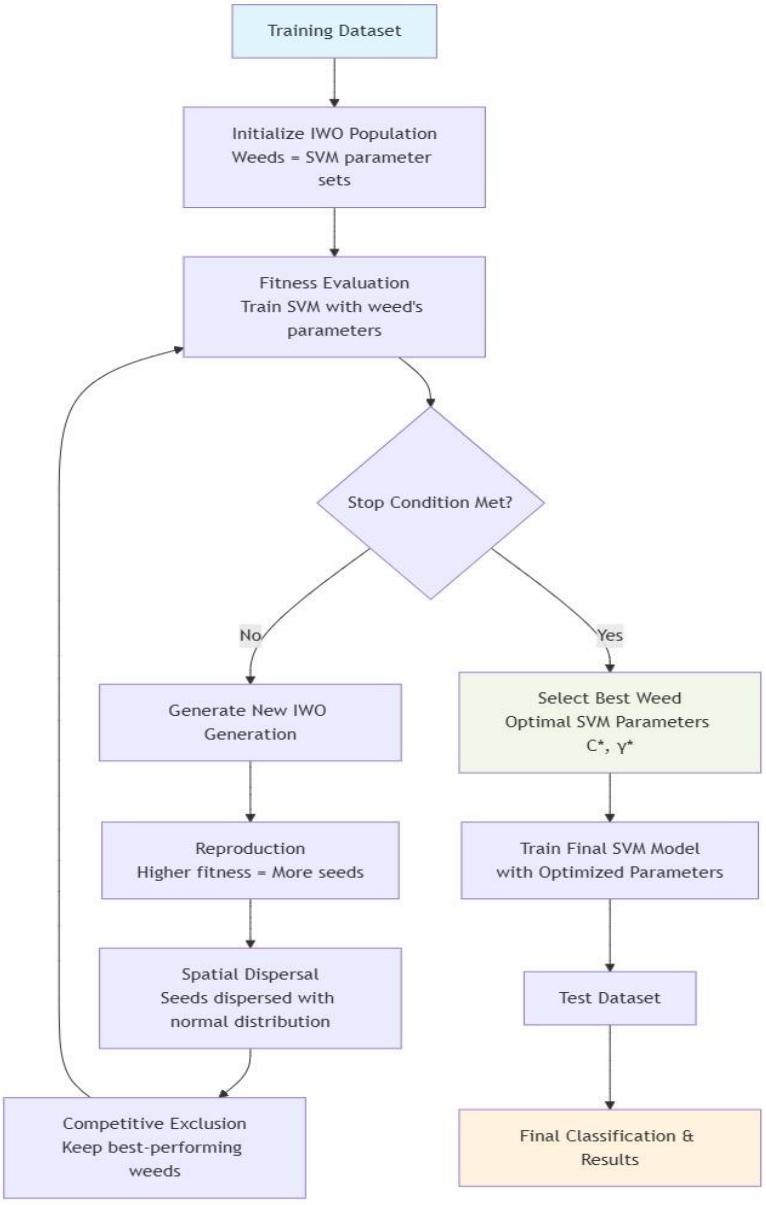


**Figure2. SVM-TLBO model structure**

### 2.2 SVM-IWO Model

The SVM-IWO (Support Vector Machine – Invasive Weed Optimization) model is a hybrid approach for predictive modeling. SVM is used for regression and classification. population produces seeds based on its fitness. From a minimum to a maximum, a plant's capacity to produce seeds varies linearly, and weeds with greater fitness yield more seeds. The production of seeds can be described as:

(3)

$$Seed_n = \frac{f - f_{min}}{f_{max} - f_{min}}(S_{max} - S_{min}) + S_{min}$$

where S_max and S_min indicate the lowest and maximum number of seeds that can be produced, f is the current fitness, f_min and f_max reflect the minimum and maximum fitness in the current population, and ⟦Seed⟧_n is the number of seeds produced. In the process of spatial dispersal, seeds are dispersed at random throughout the multidimensional search space. As the dissemination proceeds normally, seeds are kept near their parent plants. Over iterations, the standard deviation falls from σ_initial to σ_final, which is defined as:

$$\sigma_{iter} = \frac{(iter_{max} - iter)^n}{(iter_{max})^n}\left(\sigma_{initial} - \sigma_{final}\right) + \sigma_{final} \quad (4)$$

The standard deviation at the current iteration is denoted by $\sigma_{iter}$, the maximum number of iterations is indicated by $iter_{max}$, and n is a nonlinear modulation index.

Finally, competitive exclusion removes weaker seeds when the total colony size reaches its maximum ($P_{max}$). This ensures that the population evolves toward better fitness.

To assess the models' effectiveness, two widely used statistical metrics, the coefficient of determination ($R^2$) and the root mean square error (RMSE), were taken into account. The formulas for these indexes are shown below. According to these metrics, the model that performs the best is the one with the highest $R^2$ and the lowest RMSE.

(5)

$$R^2 = \left[\frac{\sum_N^{i=1}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_N^{i=1}(x_i - \bar{x})^2 . \sum_N^{i=1}(y_i - \bar{y})^2}}\right]^2$$

(6)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}}$$

where $x_i$denotes the observed values, $y_i$ refers to the computed (or estimated) values, and $\bar{x}$the represents the arithmetic mean of the observed data.

## 3. Result

Following the elimination of outliers to enhance the accuracy of the modeling process, the dataset was normalized. Thereafter, hybrid support vector machine approaches (SVM-TLBO and SVM-IWO) were employed for modeling, incorporating pH, turbidity, salinity and temperature as input variables for the Kim Tin River. The optimized parameters for the metaheuristic-SVM models are presented in Table 2. The outcomes of these models are summarized in Table 3. According to the evaluation metrics, the SVM-IWO was determined to be the most effective model. As shown in Figure 3, the optimal SVM-IWO model not only captures the variations between the observed and simulated data with high accuracy, but also clearly demonstrates the distribution and magnitude of the residual errors across the study area. As shown in Figure 4, the simulation variations of the SVM-TLBO model are clearly observable.

**Table 2. Description of Parameters Applied in the Hybrid Metaheuristic-SVM Models the Kim Tin River**

| | | | |
|---|---|---|---|
| Kernel Function | polynomial | Kernel Function | RBF |
| KernelScale | 8.8683 | Kernel Scale | 15.9051 |
| Box Constraint | 1.2040e+03 | box Constraint | 368.1100 |
| Epsilon | 11.6385 | Epsilon | 88.5487 |

**Table 3. Evaluation and Validation of Metaheuristic Models for the Kim Tin River**

| model | Input parameters | Test | |
|---|---|---|---|
| | | $R^2$ | RMSE(mg/l) |
| **SVM-IWO** | **pH, turbidity, salinity, temperature** | **0.74** | **63** |
| **SVM-TLBO** | **pH, turbidity, salinity, temperature** | **0.71** | **65** |

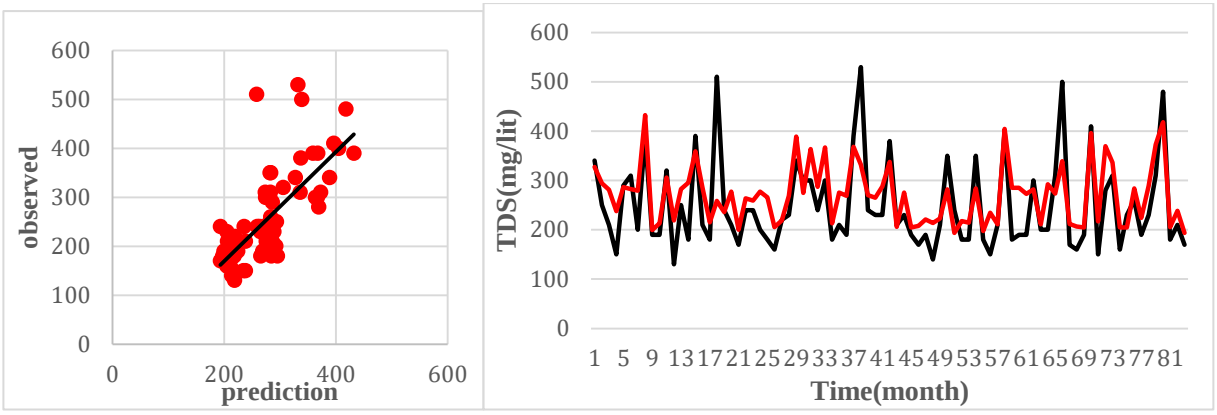


**Figure 3. Prediction Results and Validation Accuracy of SVM-IWO in Kim Tin River**

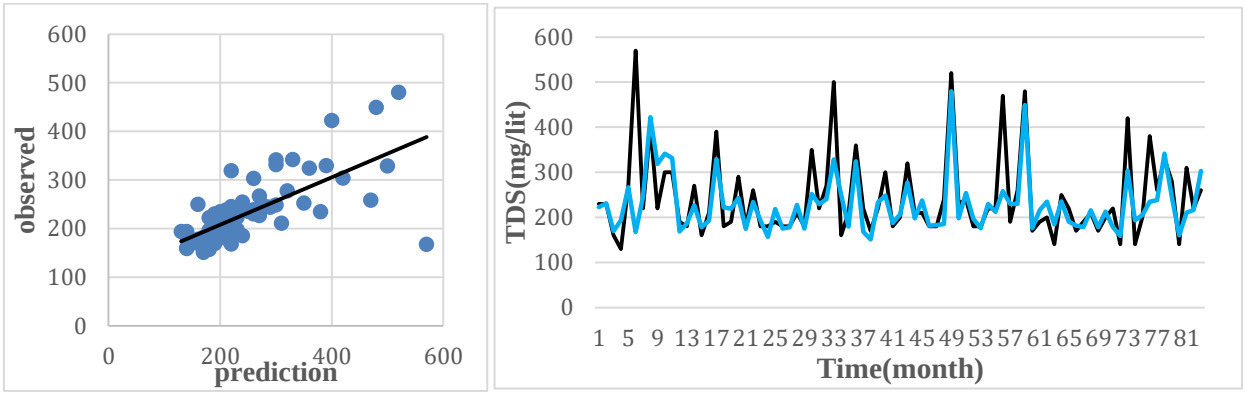


**Figure 4. Prediction Results and Validation Accuracy of SVM-TLBO in Kim Tin River**

As presented in Tables 3, the selection of the optimal solution and their comparison revealed that all two models were able to simulate the total dissolved solids (TDS) in the Kam Tin River with satisfactory accuracy and relatively low error. According to Table 3, Both models demonstrated satisfactory simulation performance, with only minor differences observed between them. Figures 3 and 4 present the same procedure, indicating that both models achieved satisfactory simulation performance.

## 4. Conclusions

This study systematically assessed the capability of hybrid SVM models for estimating total dissolved solids (TDS) in the Kam Tin River. The findings highlight that metaheuristic-optimized SVM approaches not only achieve higher predictive accuracy but also offer greater reliability compared to conventional modeling techniques. The present study evaluated the effectiveness of SVM-IWO and SVM-TLBO models in estimating total dissolved solids (TDS) in the Kam Tin River, a critical water quality parameter whose measurement is both costly and time-consuming. Monthly pH, turbidity, salinity and temperature data were employed as input variables. The results indicate that, during both training and validation phases, the SVM-IWO and SVM-TLBO models consistently outperformed the other approaches. These findings suggest that metaheuristic-based SVM models, due to their superior

predictive accuracy and reliability, can serve as a more effective and practical alternative to LS-SVM for the estimation of water quality parameters in riverine environments.

## Reference


1- Akhoni Pourhosseini, F. A., Ebrahimi, K., and Omid, M. H. (2023). Prediction of total dissolved solids based on optimization of new hybrid SVM models. *Engineering Applications of Artificial Intelligence*, *126*, 106780. https://doi.org/10.1016/j.engappai.2023.106780.

2- Barzegari Banadkooki, F., Ehteram, M., Panahi, F., Sammen, Saad Sh. Binti Othman, F., EL-Shafie, A., (2020). Estimation of total dissolved solids (TDS) using new hybrid machine learning models, Journal of Hydrology, Volume 587,2020,124989,ISSN 0022-1694,https://doi.org/10.1016/j.jhydrol.2020.124989.

3- Ghorbani, S., & Pamucar, D. (2026). Remote Sensing-Based Evaluation of Lake Area Dynamics: A Quantitative Assessment for Environmental Management in Turkey. Spectrum of Operational Research, 3(1), 352-358.

4- Goggins, W. B., Woo, J., Ho, S., Chan, E. Y., & Chau, P. H. (2012). Weather, season, and daily stroke admissions in Hong Kong. International journal of biometeorology, 56, 865-872.

5- Gulati, S., Bansal, A. & Pal, A. Estimating Total Dissolved Solids in Groundwater Using Machine Learning Models. Natural Resource Researcher 34, 1623–1644 (2025). https://doi.org/10.1007/s11053-025-10480-3.

6- Jamshidzadeh, Z., Latif, S. D., Ehteram, M., Sheikh Khozani, Z., Ahmed, A. N., Sherif, M., & El-Shafie, A. (2024). An advanced hybrid deep learning model for predicting total dissolved solids and electrical conductivity (EC) in coastal aquifers. *Environmental Sciences Europe*, *36*(1), 20. https://doi.org/10.1186/s12302-024-00850-8

7- Khatibi, R., Ghorbani, M.A. Akhoni Pourhosseini, F. (2017). Stream flow predictions using nature-inspired Firefly Algorithms and a Multiple Model strategy Directions of innovation towards next generation practices,Advanced Engineering Informatics,34,Pages 80-89,ISSN 1474-0346,https://doi.org/10.1016/j.aei.2017.10.002.

8- Liu, H. M., Grist, E. P., Xu, X. Y., Lo, H. S., Wong, A. C. Y., & Cheung, S. G. (2023). Microplastics pollution in the rivers of a metropolitan city and its estimated dependency on surrounding developed land. Science of the Total Environment, 880, 163268.

9- Mathaba, M., & Banza, J. (2023). A comprehensive review on artificial intelligence in water treatment for optimization. Clean water now and the future. Journal of Environmental Science and Health, Part A, 58(14), 1047–1060. https://doi.org/10.1080/10934529.2024.2309102

10- Memar, p., Farzin, s., (2025). Modeling and prediction of river quality parameters based on a novel hybrid machine learning model, Physics and Chemistry of the Earth, Parts A/B/C,141, Part 1,ISSNhttps://doi.org/10.1016/j.pce.2025.104067.

11- Rao, R.V., Savsani, V.J., Vakharia, D.P., 2011. Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput. Aided Des. 43, 303–315. http://dx.doi.org/10.1016/j.cad.2010.12.015.

12- Sayadi, M., Hessari, B., Montaseri, M., Naghibi, A. (2024). Enhanced TDS Modeling Using an AI Framework Integrating Grey Wolf Optimization with Kernel Extreme Learning Machine. *Water*, *16*, 2818. https://doi.org/10.3390/w16192818.

13- Singh, P., Hasija, T., Bharany, S. Tun Naeem, H.N. Chinna Rao, B. Hussen, S., Ur Rehman,A., (2025). An ensemble-driven machine learning framework for enhanced water quality classification. Discover Sustainability 6, 552. https://doi.org/10.1007/s43621-025-01467-4

14- Yildirim, F., Streimikiene, D., Bicer, A. A., Rostamzadeh, R., & Ghorbani, S. (2025). Improving monthly precipitation forecasting with GRU-LSTM model in Turkey. Acta Montanistica Slovaca., 30(1), 126-132.
15- Zheng, Y., Li, W., Fang, C., Feng, B., Zhong, Q., & Zhang, D. (2023). Investigating the impact of weather conditions on urban heat island development in the subtropical city of Hong Kong. Atmosphere, 14(2), 257.