Research Outline and Abstract of Intelligent Control-Based Water Body Health Prediction and Management

Yanjie Yin *

Department of Mechanical Engineering, Beijing Institute of Petrochemical Technology, Beijing, China

* Corresponding Author Email: 2024310422@bipt.edu.cn

Abstract. The fundamental point of the ecological balance in water bodies and the safety of drinking water is the health of water bodies. The traditional models for managing water quality are no longer effective in dealing with the environmental challenges that are now more complex due to their lagging and low efficiency. This paper tries to examine and implement the use of intelligent technologies in building an intelligent water body health management system that would encompass the prediction, early warning, and decision support. On the one hand, the system of a multi-dimensional water body health assessment that involved physical, chemical, and biological indicators was created, and the need to standardize the data was explained. Then three fundamental intelligent prediction algorithms, namely BP neural network (BPNN), fuzzy logic, and support vector machine (SVM), were thoroughly discussed, explaining their working mechanism and merits and demerits in water quality prediction. The above models were trained and validated through real monitoring data for a given reservoir using a case study. The findings revealed that the combined model indicated a high degree of benefit in the accuracy of prediction and stability. Based on this, a three-layer intelligent management system architecture including the perception layer, platform layer, and application layer was designed, achieving a management closed loop from data collection, intelligent prediction to decision support. The research results not only provide efficient and forward-looking technical means for water environment supervision but also lay a solid theoretical and practical foundation for achieving the grand goal of "smart water management".

Keywords: Water body health, Intelligent prediction, BP neural network, Fuzzy control, Support vector machine.

1. Introduction

This Water is the source of life, the foundation of production, and the basis of ecology. The well-being of water resources has a direct impact on the safety of drinking water, food production, industrial progress, and stability and equilibrium of the overall ecosystem. Nonetheless, the world is growing fast in terms of industrialization and urbanization along with the escalation of climatic change, which means increasing pressure on water bodies like never before. There is complex pollution, numerous sudden instances of pollution, eutrophication of water bodies, and decline in biodiversity, which are becoming worse. It is against this background that the conventional water management model is struggling to keep up.

The traditional management approach heavily relies on manual, fixed-point, and timed sampling along with laboratory analysis. This model has obvious limitations: Firstly, there is a significant response lag. It usually takes several hours or even days from sampling to obtaining the analysis results, making it impossible to respond promptly to instantaneous pollution incidents; Secondly, the coverage is narrow. The limited monitoring points are unable to comprehensively reflect the dynamic changes in vast water bodies; Finally, the predictive ability is weak. Management decisions are mostly based on historical data and post-event evaluations, lacking forward-looking judgments on future water quality change trends, resulting in management measures often being in a passive situation of "fixing the sheep after the sheep has been lost".

Figure 1. Industrial wastewater discharge

Figure 1 shows the discharge of industrial wastewater. Industrial wastewater discharge is one of the major sources causing water pollution, posing a serious threat to the aquatic ecosystem.

Water body management is going through radical "intelligentization" change to overcome the constriction of the classical models. The core of this transformation lies in integrating cutting-edge technologies such as the Internet of Things (IoT), big data analysis, and artificial intelligence (AI) to build an "intelligent water management" system. The introduction of intelligent control aims to achieve a fundamental paradigm shift in water body management from "passive response" to "active prediction and intervention".

The value of "intelligence" is reflected in: with the help of an online sensor network, continuous, real-time, and high-frequency monitoring parameters of water quality can be obtained; the big data platform allows collecting, cleaning, and analyzing large amounts of monitoring data; the great number of parameters of future changes in the water quality can be forecasted, and the potential hazards can be revealed through the use of artificial intelligence algorithms.

This transformation means that managers can "predict before it happens", obtaining warnings before pollution incidents occur or in the early stages of deterioration, thus gaining valuable decision-making time and taking more targeted and cost-effective intervention measures, such as adjusting upstream pollutant discharge loads, optimizing reservoir dispatching, and activating emergency treatment facilities.

Although the prospects of intelligence are vast, its implementation path still faces many technical challenges. Therefore, the core issue that this research aims to address is: How to systematically utilize intelligent technologies to construct a precise and reliable water body health prediction model, and based on this, design a closed-loop intelligent management system that can guide practical management work?

2. Core Foundation: Establishing a Scientific Water Body Health Evaluation Index System

Any accurate prediction and effective management must be based on scientific and comprehensive data. Water body health is a complex concept that cannot be measured by a single indicator. Therefore, constructing a multi-dimensional evaluation index system is the logical starting point and core foundation for all subsequent work. An effective index system should be able to comprehensively reflect the physical, chemical and biological characteristics of the water body and their interactions.

2.1. Multi-Dimensional Composition of the Index System

According to the definition of the Queensland Department of Environment, environmental indicators are "the key elements that best represent the most complex ecosystems or environmental issues, which are physical, chemical, biological or socio-economic measurements"[1]. This study draws on this framework and divides the water body health evaluation indicators into the following dimensions [2] (The seven compositions in Table 1 constitute the main body of the index system):

Table 1. Seven Composition

Physical indicators	Temperature	Turbidity	Dissolved oxygen	Chemical indicators	pH value	CO2/BOD
These are the most intuitive indicators reflecting the water body's characteristics. They mainly include	Affecting the metabolic rate of aquatic organisms and the saturation of dissolved oxygen.	Reflecting the amount of suspended matter in the water, affecting photosynthesis and the respiration of aquatic organisms.	A key factor for aquatic organisms to survive, and an important indicator for measuring the self-purification ability of the water body.	These indicators directly indicate the types and concentrations of pollutants, and are the core basis for judging the degree of water pollution. They mainly include	Reflecting the acidity or alkalinity of the water, and drastic changes can harm aquatic organisms.	Key indicators for measuring the content of organic pollutants in the water.

2.2. Construction Principles and Standardized Processing

When selecting specific indicators, a series of scientific principles must be followed, including representativeness (able to reflect key ecological processes), sensitivity (reacting sensitively to changes in environmental pressures), measurability (having mature and reliable monitoring technologies), and cost-effectiveness. In practice, trade-offs often need to be made among the effectiveness of the indicators.

Due to the significant differences in the dimensions, units, and numerical ranges of these indicators (for example, the pH value range is typically between 6 and 9, while the COD unit may be mg/L and the value can reach tens or even hundreds), directly inputting these raw data into the prediction model will lead to unstable model training processes, uneven weight distribution, and ultimately affect the prediction accuracy. Therefore, data standardization (Data Normalization) is a crucial preprocessing step before model construction.

Common standardization methods include:

Min-Max Standardization: Linearly maps the original data to the interval [0, 1] or [-1, 1]. The formula is:

$$X_{norm}=\frac{(X-X_{min})}{(X_{max}-X_{min})} \tag{1}$$

This method retains the distribution of the original data, but is very sensitive to outliers.

Z-Score Standardization: Converts the original data into a standard normal distribution with a mean of 0 and a standard deviation of 1. It is less sensitive to outliers and performs robustly in various machine learning algorithms. Relevant studies have explored the impact of different standardization techniques on the prediction effect of water quality models [3-4].

Through standardization processing, all indicators are placed on a comparable scale, ensuring that the subsequent intelligent model can fairly treat each input feature, thereby improving the stability and prediction accuracy of the model.

2.3. Support Vector Machine (SVM) Model: A Powerful Tool for Small Sample Learning

Working Principle: The Support Vector Machine (SVM) is a supervised learning model based on statistical learning theory. Its core idea is to find the optimal hyperplane in the feature space that can separate different types of samples with the maximum margin. For non-linearly separable data, SVM uses the ingenious "kernel trick" by introducing a kernel function to map the original data to a higher-dimensional feature space, thus finding a linearly separable hyperplane in the high-dimensional space, ingeniously solving the non-linear problem.

Common kernel functions include linear kernel, polynomial kernel, and the Radial Basis Function (RBF) kernel, which performs exceptionally well in water quality prediction. SVM can be used for classification tasks (such as determining whether water quality exceeds standards) and can be extended to regression tasks (SVR) for predicting specific pollutant concentrations.

Application Analysis: Compared with BPNN, SVM typically exhibits better generalization ability (i.e., performing well on unseen data) when dealing with small samples, non-linear, and high-dimensional data, and its solution is globally unique, avoiding the problem of local optimality. This renders SVM specifically beneficial in instances where there are scanty monitoring information or dimensional data. As an illustration, in a situation like that of finding the origin of a sudden pollution occurrence or the concentration of certain toxic substances, SVM can offer credible prediction outcomes due to its effective learning capability. It has been demonstrated that the SVM model can directly take measurements related to physical values as used in predicting water quality indices (WQI) with high accuracy and stable levels.

Chapter Four: Empirical Research: Case Analysis from Data to Insights

The theory is worth everything because its value is proven in practice. The entire chapter will involve a full-fledged case study to implement the above intelligent prediction model on actual data to be able to prove the real performance of the model and to obtain deeper insights into the water body management.

Ultimate Solution: Designing an Integrated Intelligent Water Quality Management System

The significance of the model of prediction is ultimately dedicated to the fact that the model is applied to the real management system. This chapter aims to integrate the aforementioned research results into an engineering manner, and design an intelligent, closed-loop water body health management system that integrates "monitoring - prediction - warning - decision-making - control", providing a feasible technical blueprint for achieving "smart water management".

2.4. BP Neural Network (BPNN) Model: A Powerful Nonlinear Fitter

Working Theory: The Backpropagation (BP) neural network is a multi-layered and feedforward network that resembles the architecture of the neuron network in the human brain. Its common architecture has an input layer, one or more hidden layers, and an output layer. The data on the standardized index of water quality is sent into the input layer and is represented as a weighted connection to the hidden layer, is subject to nonlinear transformation (via activation functions: Sigmoid or ReLU), and is sent to the output layer, finally producing the prediction result (say, the dissolved oxygen level at a given future time point).

The essence is the concept of "error backpropagation" training: The model compares the predicted value with the actual one and calculates the error; this error is being propagated backward layer by layer starting with the output layer, and depending on the error gradient, it changes the weight of every connection between one layer and another and repeats this process to the convergence process of the error output of the model. It is this mechanism that makes BPNN have strong nonlinear mapping properties, making it learn and fit the multifaceted relations between the parameters of water quality that cannot easily be expressed in a set of equations. Numerous studies have confirmed the effectiveness of BPNN in predicting river water quality [5-6].

Application Analysis: BPNN is highly suitable for regression prediction tasks with multiple variable inputs, such as using historical temperature, pH, turbidity, etc. data to predict the COD or dissolved oxygen concentration in the next 24 hours. However, BPNN also faces challenges, such as the need for empirical design of the network structure (number of hidden layers, number of neurons) and the tendency to get stuck in local optimal solutions, and sometimes the training speed is relatively slow. To overcome these shortcomings, researchers often use heuristic algorithms such as Genetic

Algorithm (GA) or Particle Swarm Optimization (PSO) to optimize the initial weights and thresholds of BPNN to improve its global search ability and convergence efficiency.

The Support Vector Machine (SVM) is a supervised learning model based on statistical learning theory.[7] Its core idea is to find the optimal hyperplane in the feature space that can separate different types of samples with the maximum margin. For non-linearly separable data, SVM uses the ingenious "kernel trick" by introducing a kernel function to map the original data to a higher-dimensional feature space, thus finding a linearly separable hyperplane in the high-dimensional space, ingeniously solving the non-linear problem.

Common kernel functions include linear kernel, polynomial kernel, and the Radial Basis Function (RBF) kernel, which performs exceptionally well in water quality prediction. SVM can be used for classification tasks (such as determining whether water quality exceeds standards) and can be extended to regression tasks (SVR) for predicting specific pollutant concentrations.

Application Analysis: Compared with BPNN, SVM typically exhibits better generalization ability (i.e., performing well on unseen data) when dealing with small samples, non-linear, and high-dimensional data, and its solution is globally unique, avoiding the problem of local optimality. This is especially beneficial to SVM in situations where there is a limited amount of monitoring data or data can be high-dimensional. As an illustration, in situations where it is necessary to establish the causes of sudden pollution or estimate the concentration of the particular toxic substances through prediction, the application of SVM could provide reasonable prediction outcomes due to its high learning capability. It has been demonstrated that the direct input of the SVM model may be the physical measurement values, which may predict the water quality indices (WQI) with outstanding accuracy and permanence [8-11].

3. System Overall Architecture

Based on the design notion of modern Internet of Things systems, the system is based on a layered, decoupled, modular structure, which is split into three layers. This design guarantees the system's scalability, flexibility, and maintenance.

One of the main devices of the intelligent management system perception layer is the buoy-type water quality monitoring station, which combines several parameter sensors and a wireless transmission module.

Sensing & Data Layer: This is the "senses" of the system. It consists of online multi-parameter water quality sensors (such as buoy stations, shore stations), meteorological stations, hydrological monitoring equipment, and unmanned aircraft/remote sensing images, deployed at key points in the reservoir. This layer is responsible for collecting real-time and continuous raw data such as water temperature, pH, DO, turbidity, and chlorophyll a, and transmitting the data securely and reliably to the cloud platform via 4G/5G or LoRa wireless communication technologies.

Platform & Model Layer: This is the "brain" of the system. It is built on cloud servers and undertakes core data processing and intelligent analysis tasks. Its main functions include: Data Hub: Responsible for receiving, storing, cleaning and managing the massive time-series data from the sensing layer.

Model Engine: Deployed and running the prediction models such as SVM and BPNN, for rolling prediction of future water quality trends.

Rule Engine: Built-in with WQI evaluation models based on fuzzy logic and expert knowledge base, used for comprehensive assessment of water quality health status and judgment of warning thresholds.

Application & Presentation Layer: This is the "window" for system interaction with users. It provides friendly user interfaces and powerful functional support to managers of different roles through web terminals and mobile APPs. Its goal is to "transform advanced data monitoring capabilities into accessible tools for professionals"[12-13].

Based on the above architecture, the system mainly implements the following three core functional modules to form a management loop:

1. Real-time Monitoring & Visualization:

Central Dashboard: Based on the GIS map, it visually displays the geographical distribution of each monitoring point and real-time water quality data. Key indicators (such as DO, COD) are presented in forms such as dynamic dashboards, line graphs, etc., and the water quality health level is marked with different colors (such as green, yellow, red).

Historical Data Query & Analysis: Users can customize the time range and indicators to query historical data of any monitoring point and conduct trend analysis, comparison, and correlation analysis to provide data support for in-depth traceability.

2. Intelligent Prediction & Alerting: Trend Prediction: The prediction model in the system's background runs automatically every hour, generating predicted curves for key water quality indicators over the next 24 hours.

Intelligent Warning: When the predicted values reach the preset warning threshold or abnormal trends are detected, the system will automatically trigger the alarm mechanism. Alarm information will be notified to relevant responsible persons through various channels such as platform pop-up windows, mobile APP push notifications, text messages, and even on-site sound and light alarms.

3. Decision Support & Coordinated Control: Intelligent decision-making suggestions: The system automatically generates emergency response recommendations based on the type and severity of the warning, combined with the built-in expert knowledge base. For example, for a predicted low oxygen level event, the system might suggest "immediately activate the No. 1 and No. 3 aeration machines and keep them running for 6 hours".

Automatic linkage control: In a higher-level system, the platform can directly link with on-site control devices (such as pumps, valves, aeration machines, and dosing devices). With authorization, the system can automatically execute control instructions according to preset rules, achieving an unmanned closed-loop management from warning to intervention.

To present the application form of the system more intuitively, the following is a simplified conceptual design of the main control dashboard. It strives to present the complex data information in the clearest and most intuitive way to the managers, enabling them to "understand the entire reservoir's health status at a glance". The Table 2 presents the conceptual design of the intelligent water body management platform - main control dashboard.

Table 2. Intelligent Water Body Management Platform

Intelligent Water Body Management Platform - Main Control Dashboard (Concept)	
GIS real-time monitoring map	Current warning event
This is an interactive map that displays the outline of the reservoir.	[High] Monitoring Point C: Predicting the next 6 hours
A blinking dot, labeled as "Monitoring Point A "(Excellent)"Monitoring Point B"	DO will be lower than 5.0mg/L
"(Good)" Monitoring Point C (Alert)". Click on the location to view detailed data.	Suggestion: Turn on the aeration machine near point C
Key indicator trend (24 hours)	Comprehensive Health Score of Reservoirs
This is a small line graph, showing the average DO, COD, and other parameters of the entire reservoir.	85 (Good)
The change of NH3-N over the past 24 hours.	

4. Conclusion

This study systematically explored the methods of water body health prediction and management based on intelligent control technology, and the main achievements and innovations are as follows:

Developed a multi-dimensional evaluation framework: Integrating physical, chemical, and biological multi-dimensional indicators, a scientific and comprehensive water body health evaluation system was constructed, and the key role of data standardization was emphasized.

Deeply analyzed the core models: Conducted in-depth analysis of the principles and application scenarios of three mainstream intelligent models, namely BP neural network, fuzzy logic, and support vector machine, clarifying their different advantages in the field of water quality prediction.

Completed empirical case studies: Through case analysis of the "Cuihu Reservoir" real data, the effectiveness and high accuracy of machine learning models such as SVM in water quality prediction tasks were verified, and the practical application value in management warnings was revealed.

Designed an integrated system solution: Proposed a hierarchical and modular intelligent water body health management system architecture, depicting a closed-loop management blueprint from data collection to intelligent decision-making, providing specific technical paths for the implementation of "Smart Water Management".

References

[1] Trach Roman, Trach Yuliia, Kiersnowska Agnieszka, Markiewicz Anna, Lendo-Siwicka Marzena, Rusakov Konstantin. A study of assessment and prediction of water quality index using fuzzy logic and ANN models. Sustainability, 2022, 14 (9): 5656.

[2] Li Jinhang, Zhao Mengdie, Han Yuping, Wei Jinhai. Analysis of water resource ecological optimization and obstacles based on the water resources–socio-economic–ecological environment model framework: a case study of the core area of the Central Plains urban agglomeration. Frontiers in Ecology and Evolution, 2023, 11: 1263601.

[3] Huang Y., Xia J., Zhang S., et al. Impacts of data quantity and quality on model calibration and validation for hydrological and water-quality simulations. Water, 2020, 12 (9): 2352.

[4] Queensland Government. Water quality indicators (factsheet collection). Brisbane: Department of Environment and Science, 2023.

[5] Islam N., Hossain M., Almutairi A., et al. Artificial ecosystem optimization with deep-learning enabled water-quality prediction and classification (AEODL-WQPC). Chemosphere, 2022, 308 (Pt 2): 136401.

[6] Shaikh Z. U. A., Korai U. A., Shaikh F. K., Talpur S. A. Smart water quality management system: a case study of Tharparkar region. VFAST Transactions on Software Engineering, 2025, 13 (1): 1 – 12.

[7] Uddin Md. Golam, Rahman M., Biswas K., et al. Data-driven evolution of water quality models: an in-depth review. Water Research, 2024, 257: 121304.2.

[8] i-manager Publications. Effects of data normalization on water quality model in a recirculatory aquaculture system using artificial neural network. Nagercoil: i-manager’s Journal on Software Engineering, 2018, 13 (1): 1 – 8.

[9] Krishnan S., Kandasamy J., Koteeswaran S., et al. Water-quality prediction: a data-driven approach exploiting machine-learning classification. Journal of Water and Climate Change, 2024, 15 (2): 431 – 447.

[10] Prabu P., Parthiban R., Nithya R., et al. Comparative analysis of machine learning models for water quality anomaly detection. Scientific Reports, 2025, 15: 15517.

[11] AIMS Press. Monitoring river water quality through predictive modeling using deep learning. AIMS Environmental Science, 2024, 11 (3): 032.

[12] Shaikh Z. U. A., Korai U. A., Shaikh F. K., Talpur S. A. Smart water quality management system: a case study of Tharparkar region. VFAST Transactions on Software Engineering, 2025, 13 (1): 1 – 12.

[13] Department of Environment and Science, Queensland Government. Water quality indicators. Brisbane: Queensland Government, 2023.