Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2021.020401
ISSN 2790-2730 Vol. 2, Issue 4: 1-12

Water Quality Evaluation Model Based on Principal Component and Particle Swarm Optimization Support Vector Machine

Dolores Blanco*

New Valley University, Egypt

**corresponding author*

Keywords: Principal Component Analysis, Particle Swarm Optimization, Support Vector Machine, Water Quality Evaluation

Abstract: Water is an important natural resource and an essential material guarantee for human survival and development. With the continuous development and progress of social economy, people's demand for water resources is higher and higher, but the current water environment is not optimistic, so the water quality evaluation and management is very urgent. In practical work, in order to ensure that the water quality reaches the standard, a variety of methods need to be used for testing, the most common of which is water sample analysis. However, due to the immature testing technology, and most of the tasks of water sample determination are completed manually, many samples are difficult to obtain accurately, resulting in large deviation of the test results, which cannot meet the current water quality monitoring requirements. Principal component analysis, particle swarm optimization (PSO) and support vector machine (SVM) models are widely used in the research field of water pollutant content as effective prediction and analysis methods, which can well link various elements in the water environment and realize reasonable judgment and control of the impact laws and change laws of different substance concentration changes. At the same time, it plays a very positive role in protecting the ecological environment. In this paper, the principal component analysis model and PSO algorithm were described in detail. Combining the SVM model and the adaptive learning mechanism of PSO, a set of online water environment quality monitoring system based on PSO was designed and developed, which provided technical support for solving the problems of the lack of representativeness and poor accuracy of traditional water quality monitoring data. Compared with the traditional single neural network, the result showed that the SVM-based evaluation model had a higher prediction effect, and the accuracy was also improved by about 6.5%, which can reflect the water pollution situation more truly and reliably.

Copyright: © 2021 by the authors. This is an Open Access article distributed under the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Water resources development is a complex system engineering, which involves water environment, water conservancy engineering and other aspects. It is characterized by strong comprehensiveness, long cycle, large investment and high risk. With the economic development and population increase, urban water consumption is gradually increasing. In addition, due to the change of people's lifestyle, industrial and agricultural production and transportation, the amount of industrial enterprises' pollution is also increasing, which causes water pollution and threatens human health. Therefore, it is necessary to strengthen the rational development and utilization of surface water resources, and take effective measures to control the total discharge of pollutants, so as to improve the water quality safety level.

Water is an important part of life on the earth, and it plays an increasingly important role in human production and life. Many scholars have studied water resources from different perspectives. Wang Xiaoyan used remote sensing technology to monitor and evaluate water quality. Based on remote sensing images, it can intuitively display the water pollution status and effectively solve the water pollution problem, which provided a new means for the rational development, utilization and management of water resources [1]. Zotou Ioanna evaluated the water quality of Mediterranean lakes based on different water quality index methods. He used the fuzzy comprehensive evaluation method and entropy method to calculate the weight of each evaluation index, and established the water quality standard model of the desalination system [2]. Abbasnia Abbas used water quality index to evaluate groundwater quality and its suitability for drinking water and irrigation water, and used multi-objective fuzzy comprehensive evaluation model to calculate the quality ranking of surface water environment in different water supply source areas [3].

Rawat K S studied the water quality index and quantity of groundwater based on the geographic information system, and established a model that focuses on water quality evaluation and comprehensively considers the comprehensive analysis results of water environment, hydrological parameters and socio-economic factors [4]. Bhutiani R used water quality index to evaluate river water quality, and used regression analysis and grey correlation analysis to comprehensively evaluate heavy metals in water body, providing scientific basis for formulating reasonable water pollution prevention policies [5]. Judran Naser Hussein used the water quality index evaluation method and the fuzzy comprehensive evaluation method to carry out the pollution treatment of the river, and established the river basin environmental quality grade index system based on the analytic hierarchy process, so as to achieve the match between the water environment status and the economic development level and population distribution [6]. Water quality assessment is the basis of river ecosystem health assessment and one of the key measures to solve the problem of water resources shortage.

Neural network models such as SVM and PSO are often used to deal with nonlinear problems, and many scholars have applied them to water quality monitoring. Paul Saili believed that freshwater fish are the main source of water pollution, and evaluated the water quality and toxicity after exposure to lead nitrate to provide a basis for the prevention and control of water pollution [7]. Leong Wei Cong used SVM and least squares SVM to predict water quality index, and established a SVM-based water quality monitoring and early warning model, so as to realize real-time dynamic monitoring of all river sections in the basin [8]. Khullar Sakshi used the deep learning method to evaluate, predict and verify the river water quality, and solved the problem that the traditional water environment quality evaluation indicators are difficult to reflect the real situation of the water body, which provided scientific basis for improving the regional economic development environment [9].

Based on the combined water quality prediction system of multi-objective optimization and improved data decomposition and integration strategy, Dong Yuqi realized real-time monitoring of the concentration and change trend of various pollutants under a single water pollution situation, and can automatically generate early warning information according to different environmental conditions [10]. Aghel B used the hybrid PSO neuro-fuzzy method to model and predict water quality parameters, and improved the global search ability by introducing adaptive weighting factors, thus realizing the dynamic planning of water quality data [11]. Alizadeh Mohamad Javad used machine learning model to study the impact of river flow on the quality of estuarine and coastal waters, and proposed a comprehensive evaluation method based on multi-scale modeling and analysis, so as to realize the overall assessment of the contribution of various hydrological parameters to water quality [12]. Water quality assessment and control is an important technical means to further improve the efficiency of urban water use, protect the safety of water resources and ensure sustainable economic development on the basis of water environment quality assessment.

In order to solve the problems of large amount of calculation, long training time and easy to fall into local extremum in general neural network algorithm, this paper constructed an automatic control system suitable for different types of water environment conditions based on SVM, which can accurately simulate the actual water pollution situation and obtain high accuracy, and can well meet the requirements of online monitoring and remote monitoring in large-scale complex systems [13]. Compared with other water quality assessment related technologies, SVM has more extensive adaptability and good scalability, especially for the processing ability of unstructured massive information.

2. Basic Model of Water Quality Assessment

2.1. General Methods for Water Quality Assessment

Water quality assessment is a basic understanding of various indicators of water, including water quantity, water quality and dissolved oxygen, so as to accurately predict the trend of water body changes in time or space. Its purpose is to provide a basis for water conservancy management departments to formulate water conservancy project construction planning. In order to ensure the reasonable and effective implementation of the water conservancy project, it is necessary to have a set of scientific and reasonable methods to carry out the water quality inspection work, and determine the quality grade of the surface water and the standard requirements that should be met in all aspects by studying the factors affecting the surface water quality and their interrelationships. Common evaluation methods include simple analysis, comprehensive analysis, quantitative analysis and qualitative analysis. The differences between them are shown in Table 1.

Table 1. Comparison of general methods of water quality assessment

Method	Advantage	Shortcoming
Simple analysis	Wide application range	Low accuracy
Comprehensive analysis	High precision	Subject to objective conditions
Quantitative analysis	Strong applicability	The calculation is large and the cost is high
Qualitative analysis	Convenient, reliable and economical	Prone to bias or deviation

Simple analysis is the most commonly used and commonly used analysis method. It adds various elements in water to a certain concentration of reagents in different proportions, and then uses a quantitative instrument to determine the content of these substances, so as to calculate the corresponding relative standard value. According to the difference between this value and other parameters, it is judged whether it conforms to relevant regulations, and finally a conclusion suitable for the actual situation of the region is drawn.

The comprehensive analysis is to conduct a comprehensive investigation of a certain type of surface water from the whole basin, and combine the hydrological characteristics to find out the causes of the problems, so as to put forward improvement measures. This kind of analysis has high accuracy, but it is often limited by objective conditions and cannot meet the needs of some special circumstances.

The quantitative analysis method is to use mathematical methods to establish a mathematical model to link the main natural factors in a certain area, such as water temperature and water level, with their corresponding related factors, such as precipitation. By using mathematical statistics, the relationship between them is obtained. This method can be used not only for monitoring the river itself, but also for other water source areas, and also for urban water supply source areas, especially for the optimal allocation of water resources in small watersheds.

Qualitative analysis is a technology that reveals the nature of phenomena from a qualitative perspective through laboratory tests. The commonly used qualitative analysis methods include univariate analysis, multivariate synthesis and multiple linear regression model. This method generally only considers the role of some specific factors, and rarely involves other variables. Therefore, the qualitative method is more convenient, reliable and economical in quantity control than the quantitative analysis method, and can better reflect the regularity of the development process of things. However, the disadvantage is that it is lack of scientificity and is prone to bias, which would affect the prediction effect. Sometimes, it would cause great losses, so it is necessary to strictly grasp its principles.

2.2. Disadvantages of Traditional Evaluation Methods

When conducting water quality survey, it is often affected by various interference factors, which cannot reflect the impact of water on the environment in different time and space, thus causing the water environment quality to be substandard. In order to improve the quality and efficiency of water quality assessment, it is necessary to study and solve these problems. However, the traditional evaluation method is difficult to reflect the water environment quality comprehensively, objectively and truly, resulting in inaccurate results. In addition, there is a lack of quantitative analysis means to analyze the process of water environment change and its characteristics from a macro perspective, which brings difficulties to improve the effect of water quality monitoring. The disadvantages of traditional evaluation methods are shown in Figure 1.

First of all, due to the limitations of human factors and external conditions, its application range is limited, and it is difficult to achieve dynamic monitoring. The traditional water environment monitoring and analysis methods have limitations and cannot reflect the actual situation well. In particular, when water pollution is deep or serious, it can only play an early warning role, but cannot really solve specific problems. In addition, the purpose of predicting the trend of water quality change cannot be achieved in different periods. For example, the past practice of using a single parameter to judge the water quality situation is no longer applicable to the current situation.

Secondly, it is difficult to meet the requirements of sustainable utilization of water resources. With the economic globalization and the development of science and technology, people’s living standards are constantly improving, and the demand for water is increasing. However, at the same time, the quality of water environment is not optimistic. Some rivers and lakes have eutrophication, which affects the safety of ecological environment and causes great harm to society. Therefore, it is necessary to strengthen river basin management and reasonably develop and utilize natural resources, so as to achieve harmonious coexistence between human and nature.

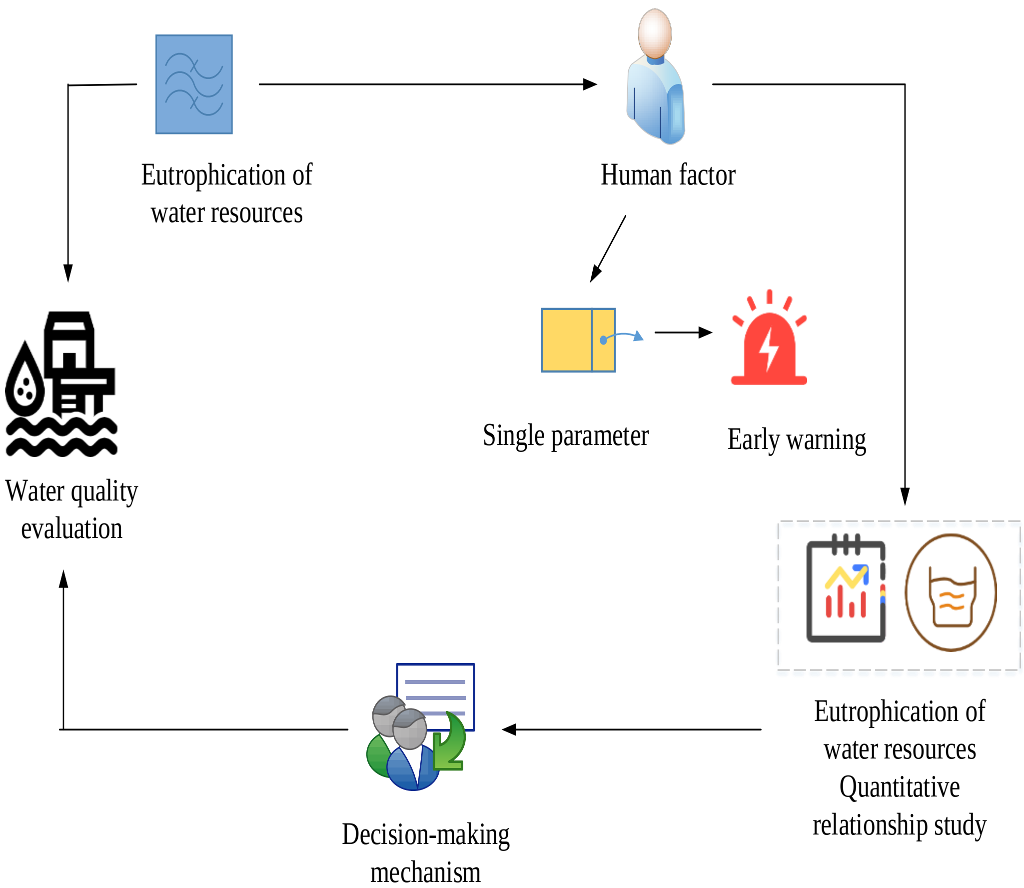


Figure 1. Disadvantages of traditional evaluation methods

Thirdly, there is no quantitative relationship between water environmental capacity and water quality indicators, and the relationship between pollutants in the water environment and human activities is not considered. In practical work, single or multiple factors are often used to describe the water environment, and the water quality parameters are not intuitive and accurate, which can not reflect the water quality status and predict the degree of pollution.

Finally, there is no scientific and effective decision-making mechanism. In the past, the water environment impact assessment was mostly qualitative and semi-quantitative comprehensive evaluation, which easily led to the subjective randomness of decision makers. Therefore, it is not conducive to scientific decision-making and supervision and implementation, and may even cause accidents, thus damaging the ecosystem and polluting the environment and other serious consequences.

2.3. Water Quality Assessment Ideas

At present, conventional methods are mainly used for water quality evaluation, such as the determination of total nitrogen content and total phosphorus content in water samples, but these data are often limited by natural conditions such as season, temperature, precipitation and so on, and it is difficult to achieve objective and fair. Another example is to take suspended solids, turbidity, chroma, etc. as the monitoring content. The chemical oxygen demand and potassium permanganate concentration are also used as test items. This traditional calculation method can no longer meet the current needs of water resources development, utilization and protection, so it is urgent to establish a comprehensive evaluation system of water environment quality suitable for the actual situation. The specific evaluation ideas are shown in Figure 2.

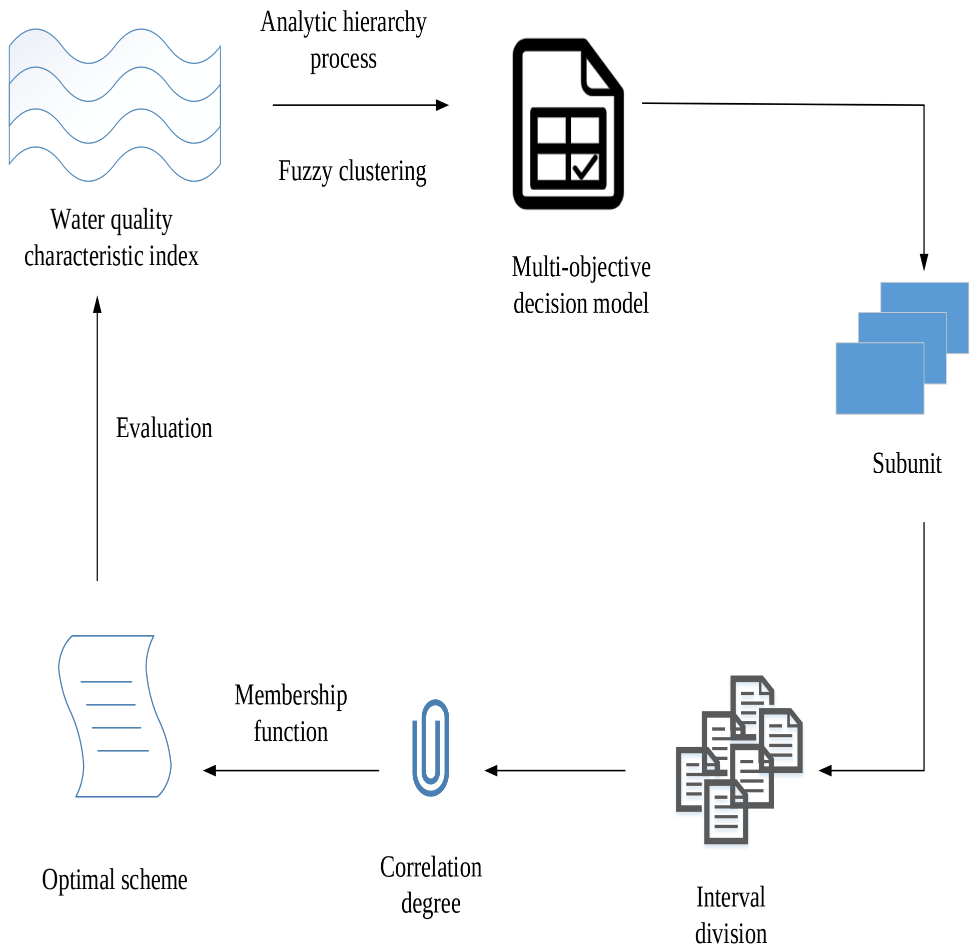


Figure 2. General ideas of water quality assessment

Firstly, a multi-objective decision-making model based on the combination of AHP and fuzzy clustering is constructed according to the water quality characteristic indicators, and the decision support system is used to assign values to each indicator respectively to obtain the corresponding weights of each sub-unit. After that, the whole area is divided into several sub-units according to the degree of correlation between each sub-block. After determining the correlation degree of each segment, the corresponding weighting coefficient is calculated to reflect the correlation between different subunits and the difference degree of influencing factors. Combined with the theory of membership function, using the principle of fuzzy mathematics, the optimal scheme is selected in order according to the sequencing principle, and finally the evaluation result of the overall water quality grade of the basin is obtained.

3. Common Technologies for Water Quality Monitoring

Water quality monitoring technology is a multidisciplinary and comprehensive application based on environmental science. Common auxiliary means include principal component analysis, PSO and SVM model, which have good prediction effect on the concentration distribution of pollutants in water [14].

3.1. Principal Component Analysis

It is supposed that there are variables x and y, and the correlation coefficient matrix between them is:

$$R=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1i} \\ a_{21} & a_{22} & \dots & a_{2i} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ii} \end{bmatrix} \tag{1}$$

Among them, a_{ii} is the correlation coefficient of variables x and y, and $a_{xy}=a_{yx}$; its definition formula is:

$$a_{xy}=\frac{\sum_{n=1}^{N}(x_{nu}-\overline{x_u})(x_{nv}-\overline{x_v})}{\sqrt{\sum_{n=1}^{N}(x_{nu}-\overline{x_u})^2\sum_{n=1}^{N}(x_{nv}-\overline{x_v})^2}} \tag{2}$$

The characteristic equation can be obtained by solving the eigenvalues, and the eigenvalues and eigenvectors can be obtained, so the appropriate m($m \le i$) principal components are selected as the new index.

3.2. Particle Swarm Optimization

It is assumed that the position and velocity of each particle in the initial particle mass are g and v respectively, then there are:

$$v(T+1)=v(T)+a_1b_1(\lambda_1-g(T))+a_2b_2(\lambda_2-g(T)) \tag{3}$$

$$g(T+1)=g(T)+v(T+1) \tag{4}$$

Among them, a_1 and a_2 are acceleration coefficients; b_1 and b_2 are random numbers between 0 and 1.

3.3. Support Vector Machine

It is supposed there is a discriminant function and a classification surface equation in the high-dimensional space. By normalizing the function and requiring the classification face to correctly classify all samples, that is, to meet the relationship:

$$f(ax+b)\ge 1 \tag{5}$$

Among them, x is the input vector, and f is the classification label.

The expression for finding the optimal classification surface is:

$$\min\frac{a^2}{2}\,f(ax+b)\ge 1 \tag{6}$$

Among them, $\frac{a^2}{2}$ is the classification interval.

3.4. Establishment of Water Quality Assessment Model

Some problems existing in the traditional water quality assessment, such as single index and complicated calculation, have become the main factors restricting its popularization and application. Moreover, such methods often cannot accurately reflect the water environment quality when applied, so it is necessary to improve the current water quality assessment system [15]. In view of the above problems, this paper proposed a comprehensive evaluation model based on the combination of PSO and SVM, as shown in Figure 3.

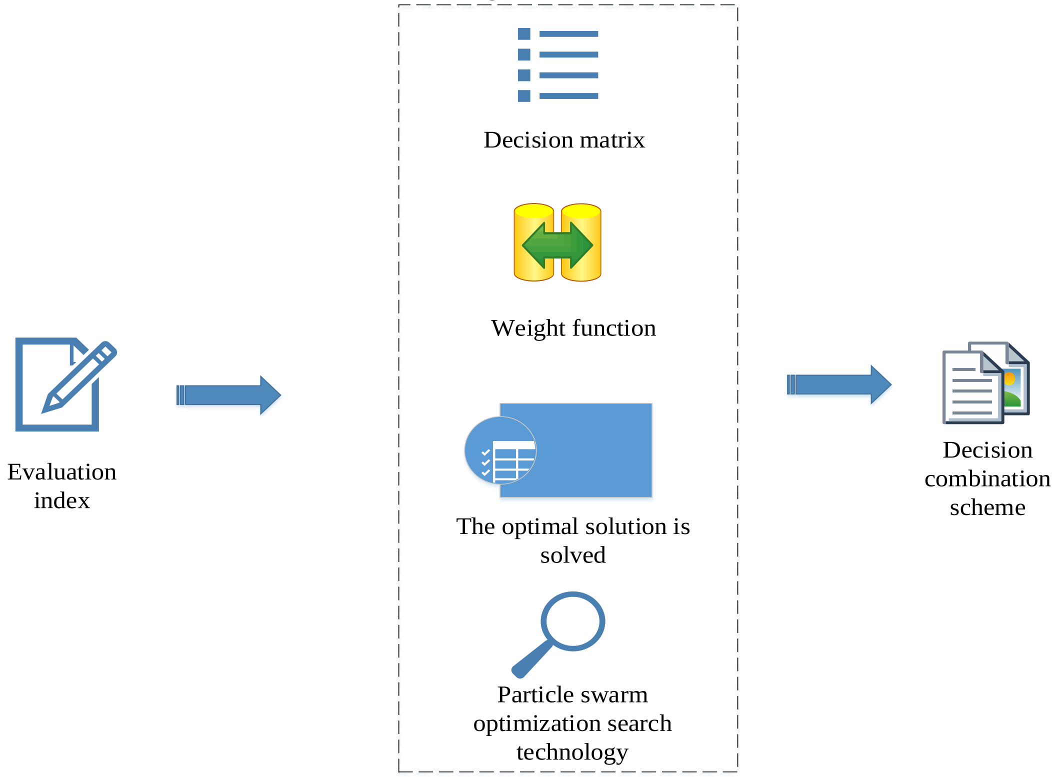


Figure 3. Establishment of water quality evaluation model

First of all, the water quality evaluation indicators include total nitrogen, total phosphorus, ammonia nitrogen, dissolved oxygen, and chemical oxygen demand. These indicators are used to establish a fuzzy multi-attribute decision-making matrix applicable to various water quality conditions. Secondly, SVM is used to construct the weight function, and genetic algorithm is introduced to solve the optimal solution. Finally, a new decision combination scheme is constructed by combining the algorithm with PSO search technology, which makes this combination scheme easier to implement and realize, thus improving the efficiency and accuracy of the integration process. This model can effectively solve the shortcomings that the existing evaluation model cannot reflect the relationship between the quality of water and the interaction between the relevant factors is not obvious. This is convenient to guide the selection of river control schemes and provide basis for improving the water environment quality of the basin. At the same time, it can also provide decision-making basis for the water resources management department to formulate the river water dispatching plan and relevant policy decisions.

4. Water Quality Evaluation Experiment and Results

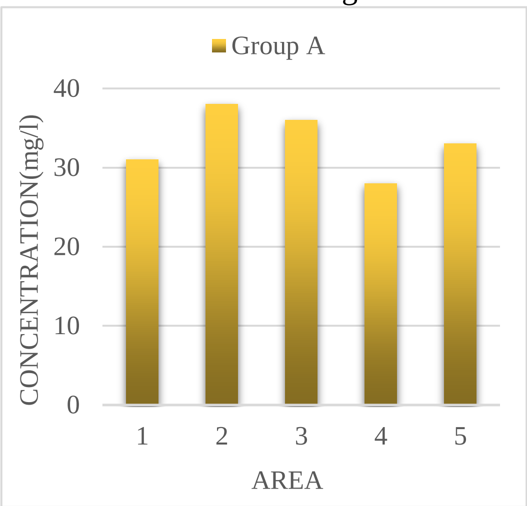
4.1. Experimental Method

Taking a river basin as the research object, two groups were assigned to monitor the water quality in five areas of the river basin. It is known that Group A adopted the traditional neural network evaluation method, and Group B adopted the SVM-based evaluation method. With the change and accuracy of chemical oxygen demand as the evaluation index, the monitoring effect and water quality change of the two groups were compared.

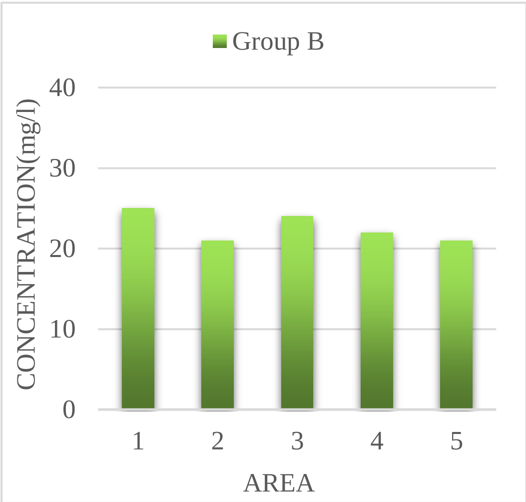
4.2. Data

4.2.1. Chemical Oxygen Demand

The detection principle of chemical oxygen demand is to use potassium permanganate to oxidize and decompose oxidizable substances in water, such as organic matter, nitrite, ferrous salt, sulfide, etc. After that, the oxygen consumption was calculated according to the amount of residual oxidant, and the change trend of chemical oxygen demand of the two groups was calculated. The higher the value of chemical oxygen demand, the more serious the water pollution, and the worse the control effect. The results are shown in Figure 4.



a. Group A concentration change



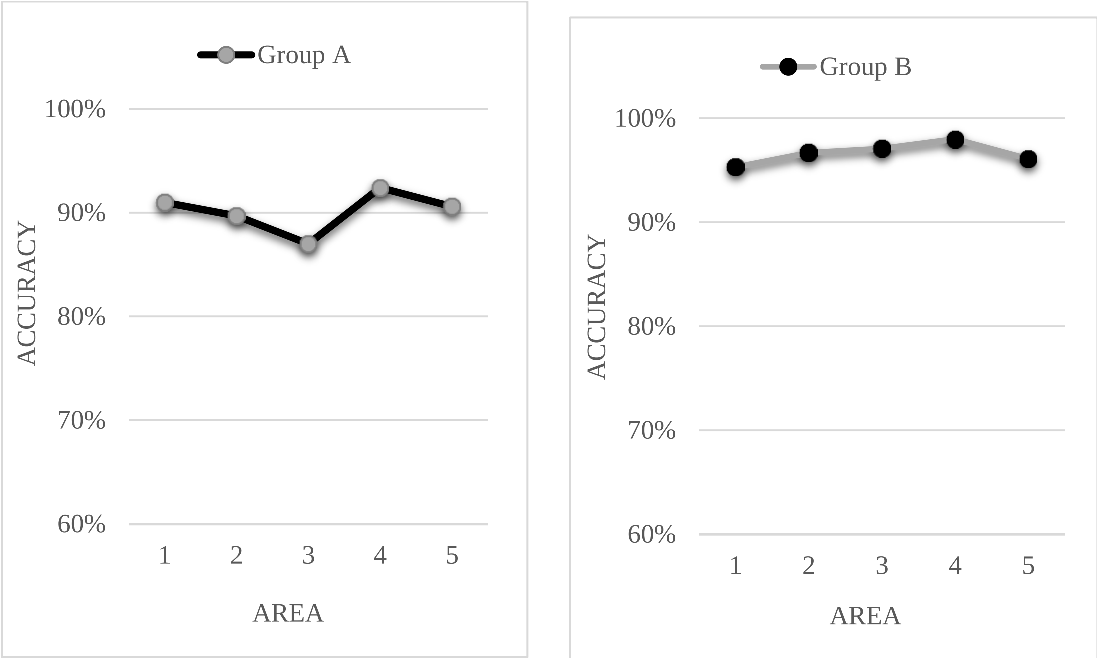
b. Group B concentration change

Figure 4. Comparison of changes in chemical oxygen demand between the two groups

Figure 4a shows the concentration change of Group A, and Figure 4b shows the concentration change of Group B. From Figure 4, the concentration changes in the five regions monitored can be seen. The concentration range of group A was 25~40mg/l, and that of group B was 20~25mg/l. After the prevention and treatment of neural network and SVM model, the overall concentration range of group A was higher, but the distribution curve was not smooth; the floating range of group B curve was smaller. Therefore, it is feasible to predict by SVM model.

4.2.2. Accuracy

The accuracy of the two groups of evaluation results was compared. The results are shown in Figure 5.



a. Group A accuracy *b. Group B accuracy*

Figure 5. Comparison of accuracy of evaluation results between the two groups

Figure 5a shows the accuracy of Group A and Figure 5b shows the accuracy of Group B. It can be clearly seen that the curve of Group B was distributed above Group A; the curve had almost no fluctuation and the trend was relatively flat, which showed that the evaluation model based on SVM had high stability. On the other hand, in Group A, the curve experienced a decline first, then a rise, and finally became stable, with the accuracy floating around 90%. After calculation, the average accuracy of Group A was about 90.14%, and that of Group B was about 96.64%, which was about 6.5% higher than that of Group A.

5. Conclusion

The water quality assessment model is the basis of water quality management and an important standard to measure the quality of water environment. Its effectiveness is directly related to the sustainable use of water resources and environmental resources. Applying SVM and PSO algorithms to hydrological forecasting can achieve good adaptability to solving complex nonlinear problems. The comprehensive evaluation model designed in this paper can effectively solve the problem that the conventional neural network model is difficult to adapt to the complex environment, and can better meet the actual engineering requirements, which is of great significance to promote the water resources management department to strengthen the prevention and control of water pollution.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Xiaoyan Wang, Yang Wu. Water quality monitoring and evaluation using remote sensing techniques in China: A systematic review. Ecosystem Health and Sustainability. (2019) 5(1): 47-56. https://doi.org/10.1080/20964129.2019.1571443

[2] Zotou Ioanna, Vassilios A. Tsihrintzis, Georgios D. Gikas. Water quality evaluation of a lacustrine water body in the Mediterranean based on different water quality index (WQI) methodologies. Journal of Environmental Science and Health, Part A. (2020) 55(5): 537-548. https://doi.org/10.1080/10934529.2019.1710956

[3] Abbasnia Abbas. Evaluation of groundwater quality using water quality index and its suitability for assessing water for drinking and irrigation purposes: Case study of Sistan and Baluchistan province (Iran). Human and Ecological Risk Assessment: An International Journal. (2019) 25(4): 988-1005. https://doi.org/10.1080/10807039.2018.1458596

[4] Rawat K. S., Sudhir Kumar Singh. Water Quality Indices and GIS-based evaluation of a decadal groundwater quality. Geology, Ecology, and Landscapes. (2018) 2(4): 240-255. https://doi.org/10.1080/24749508.2018.1452462

[5] Bhutiani R. Evaluation of water quality of River Malin using water quality index (WQI) at Najibabad, Bijnor (UP) India. Environment Conservation Journal. (2018) 19(1-2): 191-201. https://doi.org/10.36953/ECJ.2018.191228

[6] [6] Judran Naser Hussein, Ajeet Kumar. Evaluation of water quality of Al-Gharraf River using the water quality index (WQI). Modeling earth systems and environment. (2020) 6(3): 1581-1588. https://doi.org/10.1007/s40808-020-00775-0

[7] Paul Saili. Evaluation of water quality and toxicity after exposure of lead nitrate in fresh water fish, major source of water pollution. The Egyptian Journal of Aquatic Research. (2019) 45(4): 345-351. https://doi.org/10.1016/j.ejar.2019.09.001

[8] Leong Wei Cong. Prediction of water quality index (WQI) using support vector machine (SVM) and least square-support vector machine (LS-SVM). International Journal of River Basin Management. (2021) 19(2): 149-156. https://doi.org/10.1080/15715124.2019.1628030

[9] Khullar Sakshi, Nanhey Singh. Water quality assessment of a river using deep learning Bi-LSTM methodology: forecasting and validation. Environmental Science and Pollution Research. (2021) 29(9): 12875-12889. https://doi.org/10.1007/s11356-021-13875-w

[10] Yuqi Dong. Combined water quality forecasting system based on multiobjective optimization and improved data decomposition integration strategy. Journal of Forecasting. (2021) 42(2): 260-287. https://doi.org/10.1002/for.2905

[11] Aghel B., A. Rezaei, M. Mohadesi. Modeling and prediction of water quality parameters using a hybrid particle swarm optimization-neural fuzzy approach. International Journal of Environmental Science and Technology. (2019) 16(8): 4823-4832. https://doi.org/10.1007/s13762-018-1896-3

[12] Alizadeh Mohamad Javad. Effect of river flow on the quality of estuarine and coastal waters using machine learning models. Engineering Applications of Computational Fluid Mechanics. (2018) 12(1): 810-823. https://doi.org/10.1080/19942060.2018.1528480

[13] Bisht Anil Kumar. Artificial neural network based water quality forecasting model for ganga river. International Journal of Engineering and Advanced Technology. (2019) 8(6): 2778-2785. https://doi.org/10.35940/ijeat.F8841.088619

[14] Solangi Ghulam Shabir. Groundwater quality evaluation using the water quality index (WQI), the synthetic pollution index (SPI), and geospatial tools: a case study of Sujawal district, Pakistan. Human and Ecological Risk Assessment: An International Journal. (2020) 26(6): 1529-1549. https://doi.org/10.1080/10807039.2019.1588099

[15] Gad Ahmed G. Particle swarm optimization algorithm and its applications: a systematic review. Archives of computational methods in engineering. (2021) 29(5): 2531-2561. https://doi.org/10.1007/s11831-021-09694-4