Scholar Publishing Group
Water Pollution Prevention and Control Project
https://doi.org/10.38007/WPPCP.2022.030102
ISSN 2790-2730 Vol. 3, Issue 1: 12-21

Support Vector Machine Algorithm in Water Quality Prediction

Georgy V. Moiseev*

Institute of IT & Computer Science, Afghanistan

**corresponding author*

Keywords: Water Quality Prediction, Support Vector Machines, Neural Networks, Water Resources

Abstract: Water quality assessment and prediction is an essential part of aquatic research. Its purpose is to accurately grasp the situation of water quality and pollutants and predict the development trend in the future, which is a basic task of managing, protecting and maintaining the water environment. The increasingly serious water pollution has destroyed the ecological environment, affected people's life and health, and seriously restricted social and economic development. Based on this, it is necessary to predict the water quality. Therefore, this paper designed support vector machine (SVM) algorithm, designed water quality prediction model, and utilized it to water quality prediction. At the same time, experiments were designed to compare the SVM algorithm with the traditional neural network algorithm. The average concentration value of permanganate index and ammonia nitrogen concentration of water quality parameters were selected for prediction and analysis, and finally a feasible conclusion was reached. Compared with the traditional neural network algorithm, SVM algorithm could greatly improve the prediction effect of the average concentration of permanganate index and ammonia nitrogen concentration. The error between the predicted average concentration of permanganate index and the actual value was less than 3%, which was conducive to improving the prediction effect of water quality. The research results showed that SVM algorithm had good prediction effect in water quality prediction and could expand the application range.

1. Introduction

Water is closely related to people's life. Human survival and development cannot be separated from water resources, and the survival and development of the earth cannot be separated from water resources. The prediction of water resources quality can not only effectively manage water resources and avoid pollution, but also help people to make the best use of water resources. The prediction of water resources quality can greatly improve the utilization effect of water resources, and the water quality management and analysis can be carried out according to the prediction results. Therefore, there is a need for water quality prediction.

Water quality prediction is widely used at present, and some scholars have made great achievements in this field. Bisht Anil Kumar used artificial neural network (ANN) to establish the water quality prediction model of Ganga River [1]. Yang Huanhai established a water quality prediction model for aquaculture using multi-scale decomposition [2]. Solangi Ghulam Shabir assessed and mapped the groundwater quality in Sujawar district of coastal region of Pakistan [3]. Haghiabi Amir Hamzeh used machine learning method to predict water quality [4]. Ayaz Muhammad predicted the heavy metal pollution in the coastal sea waters of Karachi Port Area through neural network method [5]. Barzegar Rahim developed two independent deep learning models to predict the two water quality variables of Little Prespa Lake in Greece [6]. Hassan Md Mehedi used machine learning algorithm to efficiently predict water quality index [7]. In the analysis of water quality prediction, machine learning algorithm is often used, and SVM algorithm is rarely used.

SVM algorithm is also applied in water quality analysis. Leong Wei Cong studied the prediction of water quality index based on SVM and least squares SVM [8]. Meshram Sarita Gajbhiye developed four machine learning models including SVM models to predict daily sediment production [9]. Wang Lei proposed a spatially distributed SVM system for estimating shallow water bathymetry from optical satellite images [10]. Shamshirband Shahaboddin used the nested grid numerical model and the comparison between ANN, extreme learning and machine learning models of SVM to predict important wave heights [11]. Seifi Akram proposed a mixed model of least squares support vector gamma test to estimate the water evaporation under drought conditions at Zahran station in Iran [12]. No one has built SVM water quality prediction model to analyze water quality prediction.

To enhance the accuracy of water quality prediction, this paper used SVM algorithm to predict water quality. First, the importance of water quality prediction was introduced, and then the traditional water quality prediction algorithm was analyzed. Then, the SVM algorithm and its parameter importance were introduced. According to SVM algorithm, water quality prediction model was designed and applied to water quality prediction. Finally, compared with the traditional neural network algorithm, SVM algorithm could greatly enhance the accuracy of water quality prediction and achieve the expected results. Compared with other people’s experiments, this paper selected the average concentration of permanganate index and ammonia nitrogen concentration as water quality parameters for analysis.

2. Importance of Water Quality Prediction

Predicting water quality is a prerequisite for safe water use [13]. The quality of drinking water is directly related to human health [14]. Therefore, each production unit of water supply must carry out rigorous prediction and sampling. Only qualified water can enter the water supply network. Drinking water standards have also been set higher standards. Only a strict prediction of drinking water can ensure its safety. Common water quality prediction includes water quality prediction of water source and water supply network, as shown in Figure 1.

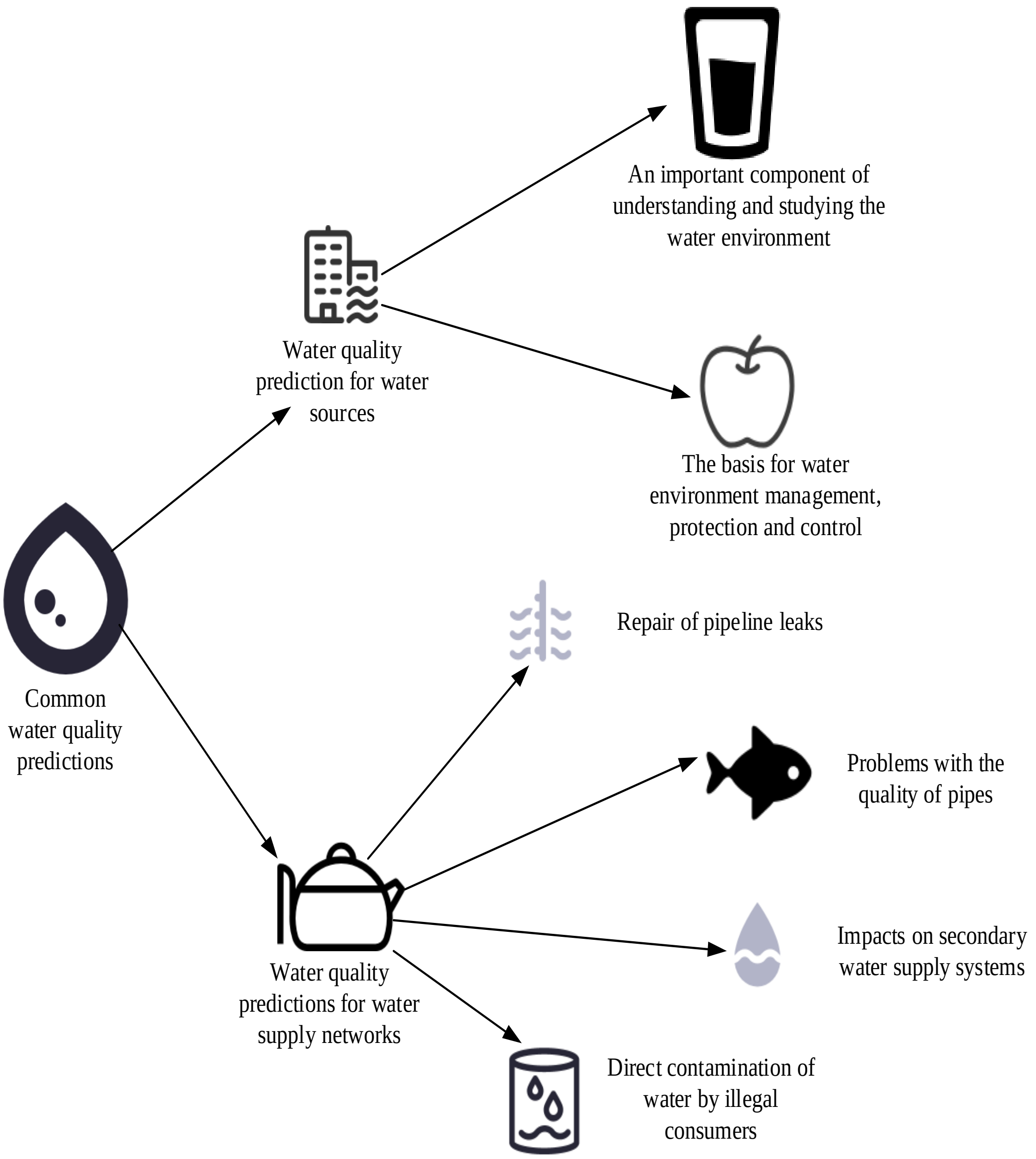


Figure 1. Common water quality predictions

2.1. Prediction of Water Quality

The quality of water source is the source of water quality [15]. In the production process of the waterworks, the change of water quality must be closely predicted, so that the quality characteristics of raw water can be understood timely and accurately, and the subsequent treatment stage can be adjusted timely and smoothly [16]. Water pollution prediction has a good application value for the treatment and protection of water sources. The important components of the water environment can be analyzed to avoid the adverse effects of water pollution on people. The water quality model can quantitatively simulate and predict the water environment quality, which provides technical support for water pollution prevention and control planning and other related management and planning tasks, and provides effective support for water pollution control and water source management of lakes and reservoirs. Most groundwater also needs special treatment to meet the drinking water or industrial water standards.

2.2. Water Quality Prediction of Water Supply Network

The quality of drinking water discharged into the water supply network is strictly controlled at all stages of production [17-18]. However, this does not mean that the drinking water supplied to consumers can fully meet the sanitary requirements of drinking water, because drinking water as a special commodity has its own characteristics. After leaving the factory, drinking water is delivered to consumers through a series of pipes. During the continuous and liquid transportation of drinking water, there are many factors that affect the secondary pollution of water quality, such as the repair of pipeline leakage, the problem of pipeline quality, the impact on the secondary water supply system, and the direct pollution of illegal consumer water.

3. Traditional Water Quality Prediction Algorithm

There are more and more changing factors affecting the water environment, and the changes are also increasing. The uncertainty of the water environment is also increasing gradually [19-20]. From the current research situation, the estimation range obtained by the fuzzy estimation method using the linear weighted average method is prone to outliers, errors, spikes and other phenomena, and there are problems of inaccurate estimation or inconsistent results of water quality categories, which have little effect. In the application of fuzzy theory for comprehensive evaluation of water quality, it is necessary to solve the problem of fair weighting and comparability, which needs further research. The advantages of fuzzy method for evaluation and prediction are simplicity and comparability, while the disadvantages are low resolution. ANN is good at dealing with nonlinear dependence. However, previous studies have shown that neural networks have some shortcomings that are difficult to overcome, such as network structure is difficult to determine, local optimization, easy to readjust, and poor generalization ability. Therefore, it limits the application effect of neural network to a certain extent.

Effective data mining tools include statistical pattern recognition, linear or nonlinear regression and ANN. Due to the rapid development of software and hardware technology, these data mining tools are now widely used. There is a “small sample problem”, because many practical applications only have a small number of known samples, while traditional pattern recognition or ANN methods require a large number of training samples. Before the advent of SVM technology, the solution of the “small sample problem”, that is, a small number of samples produce a highly generalized model, is a problem in pattern recognition research.

4. SVM Algorithm

Compared with traditional neural network training methods, SVM is on the basis of the principle of structural risk minimization. By addressing the quadratic programming issue, the global optimal solution is obtained, effectively solving the problems of model selection and relearning, nonlinear and dimensional catastrophes, and local minima.

SVM realizes the mapping of nonlinear distributed sample bending space to high-dimensional feature space. The idea that specific nonlinear mapping methods are used to divide them linearly. Such high-dimensional feature space is used to construct an optimal classification hypersurface that can be classified. SVM essentially solves the secondary programming problem, so it has the advantages of high versatility, easy learning and no local minimum.

Assuming that the given training sample is:

$$\{(a_1,b_1),(a_2,b_2),\Lambda,(a_n,b_n)\}\subset R_n\times R \tag{1}$$

The soft variable β_j, β'_j is introduced to form the error-proof penalty coefficient D and constant γ. Let the regression function have the following form:

$$f(a) = (\omega \bullet \alpha(a)) + b \tag{2}$$

Then, the form of SVM model to solve this problem is:

$$\min \frac{1}{2}\|\omega\|^2 + D\sum_{j=1}^{n}(\beta_j + \beta'_j) \tag{3}$$

Lagrange function is introduced and its dual form is obtained:

$$\sum_{j=1}^{n} b_j(x_j - x'_j) - \gamma\sum_{j=1}^{n}(x_j + x'_j) - \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{n}(x_j - x'_j)(x_i - x'_i)Z(a_j, a_i) \tag{4}$$

The final regression function can be obtained:

$$f(a) = \sum_{j=1}^{n}(x_j - x'_j)Z(a_j \bullet a) + b \tag{5}$$

5. Construction of SVM Parameters and SVM Model

5.1. SVM Parameters

When using SVM algorithm for regression and prediction, the selection of kernel function and the optimization of parameters are important criteria to measure the accuracy of the model. In the SVM estimation task, the selection of error penetration coefficient, kernel function and their parameters have an essential impact on classification accuracy. The kernel parameters include the order of the polynomial kernel function, the order of the radial kernel function, and the width and offset of the sigmoid kernel function.

The parameter is a penalty factor for the variance of misclassified samples. Its function is to adjust the ratio of empirical risk and confidence interval of the training machine to achieve the best generalization effect. A balance must be found between reducing the number of misclassified samples (reducing empirical risk) and increasing the classification interval (reducing confidence interval). The low value of the parameter means low training error penalty and less complex training machine.

5.2. Construction of SVM Model for Water Quality Prediction

The specific steps to establish the SVM model for water quality prediction are shown in Figure 2.

Step 1: The learning samples are collected and pretreated. Step 2: Kernel function and penalty coefficient are selected. In this paper, radial kernel function and kernel parameters are selected. Step 3: The secondary SVM optimization problem is constructed and the algorithm is decomposed. The minimum sequence algorithm is used in this work. Step 4: The values of a and b are obtained and inserted into the equation to obtain the best classification function, namely the classification model of SVM. Step 5: The test sample is inserted into the model to get the test results.

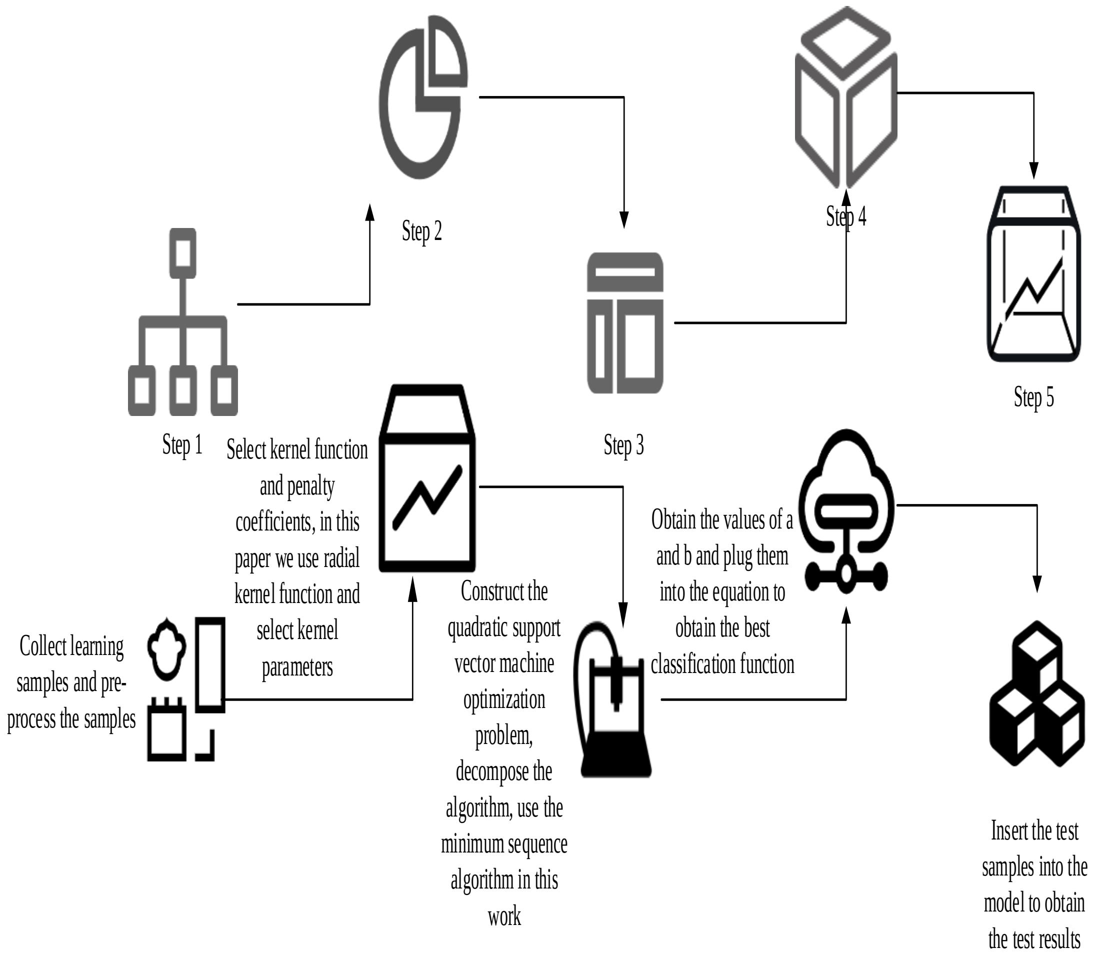


Figure 2. Construction of the SVM model for water quality prediction

6. Prediction Effect of Water Quality

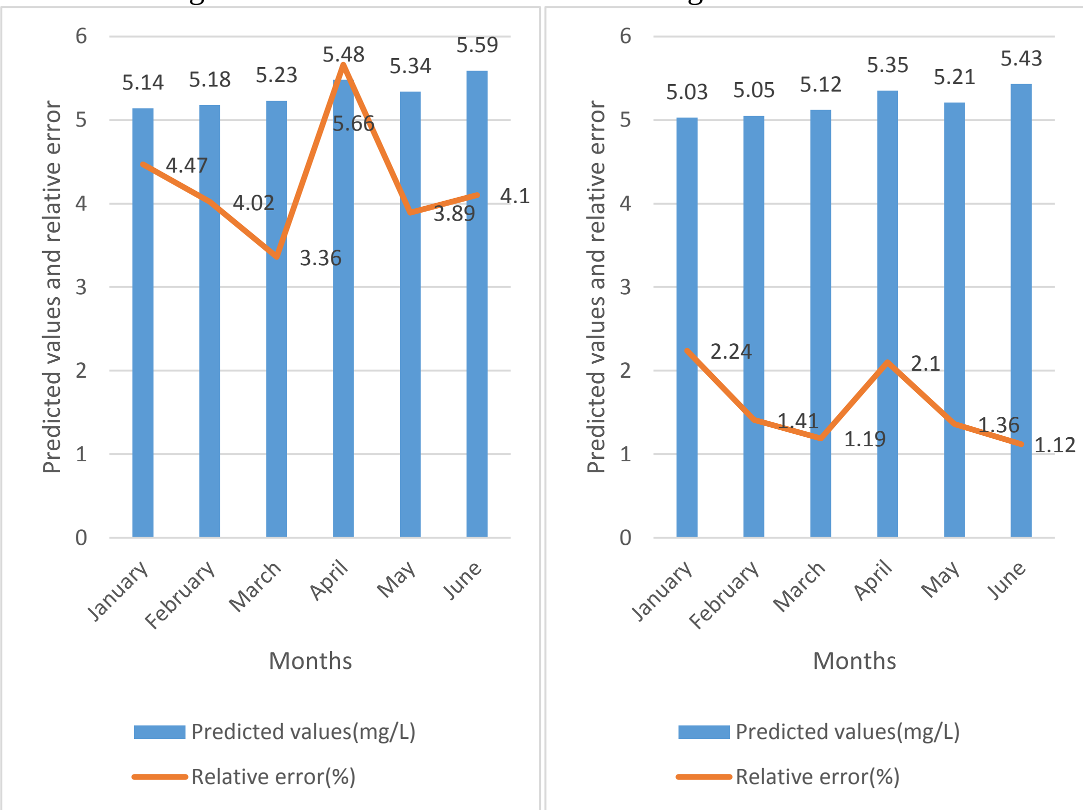
To analyze the prediction effect of the SVM model proposed in this article, this article selected the average concentration of permanganate index (unit: mg/L) and ammonia nitrogen concentration (unit: mg/L) of water quality parameters for prediction and analysis, and selected a water source. The average concentration of permanganate index and the measured value of ammonia nitrogen concentration from January to June 2022 are recorded in Table 1.

Table 1. Actual measured values of water quality parameters at the water source

Months	Average permanganate index concentration(mg/L)	Ammonia nitrogen concentration(mg/L)
January	4.92	0.65
February	4.98	0.71
March	5.06	0.83
April	5.24	1.18
May	5.14	1.49
June	5.37	1.85

6.1. Average Concentration of Permanganate Index

The average concentration value of permanganate index predicted by traditional neural network algorithm and SVM algorithm and its error are recorded in Figure 3.



A. Average permanganate index concentration and relative error predicted by neural network algorithm

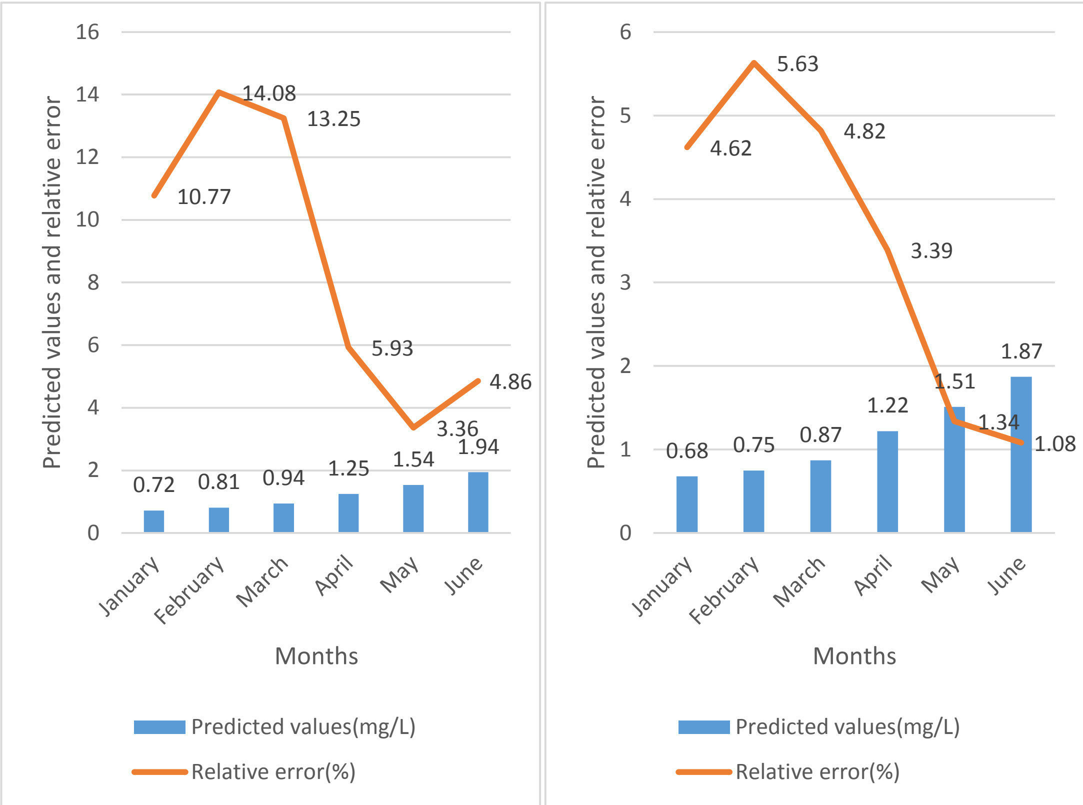
B. Average permanganate index concentration and relative error predicted by support vector machine algorithm

Figure 3. Average permanganese index concentration and relative error predicted by different algorithms

In Figure 3, A represents the average concentration and relative error of permanganate index predicted by neural network algorithm, and B represents the average concentration and relative error of permanganate index predicted by SVM algorithm. The error between the predicted average concentration value of permanganate index and the actual value from January to June 2022 using neural network algorithm was more than 3%. However, the error between the average concentration of permanganate index predicted by SVM algorithm and the actual value from January to June 2022 was less than 3%. It shows that SVM algorithm has greatly improved the accuracy of water quality prediction compared with neural network algorithm.

6.2. Ammonia Nitrogen Concentration

The ammonia nitrogen concentration value and its error predicted by traditional neural network algorithm and SVM algorithm are recorded in Figure 4.



A. Ammonia nitrogen concentration and relative error using neural network algorithm

B. Ammonia nitrogen concentration and relative error using support vector machine algorithm

Figure 4. Ammonia nitrogen concentration and relative error predicted by different algorithms

In Figure 4, A represents the ammonia nitrogen concentration and relative error predicted by neural network algorithm, and B represents the ammonia nitrogen concentration and relative error predicted by SVM algorithm. When two algorithms were used to predict the concentration of ammonia nitrogen, it was found that the two algorithms had a large error when predicting the lower concentration of ammonia nitrogen, and a small error when predicting the higher concentration of ammonia nitrogen. However, in general, the ammonia nitrogen concentration predicted by SVM algorithm was close to the actual value, and the relative error was also small. It shows that SVM algorithm can greatly improve the accuracy of ammonia nitrogen concentration prediction.

When using the neural network algorithm, the predicted ammonia nitrogen concentration in January was 0.72 mg/L, and the relative error was 10.77%. The predicted ammonia nitrogen concentration in February was 0.81mg/L, with a relative error of 14.08%. The predicted ammonia nitrogen concentration in March was 0.94mg/L, with a relative error of 13.25%. The predicted ammonia nitrogen concentration in April was 1.25 mg/L, and the relative error was 5.93%. The predicted ammonia nitrogen concentration in May was 1.54 mg/L, and the relative error was 3.36%. The predicted ammonia nitrogen concentration in June was 1.94mg/L, with a relative error of 4.86%.

The predicted ammonia nitrogen concentration in January was 0.68 mg/L and the relative error was 4.62% when using and SVM algorithm. In February, the predicted ammonia concentration was 0.75 mg/L, and the relative error was 5.63%. The predicted ammonia concentration in March was 0.87 mg/L, and the relative error was 4.82%. In April, the predicted ammonia concentration was

1.22 mg/L, and the relative error was 3.39%. In May, the predicted ammonia concentration was 1.51 mg/L, and the relative error was 1.34%. The predicted ammonia concentration in June was 1.87 mg/L, and the relative error was 1.08%. In comparison, SVM algorithm can greatly improve the prediction accuracy of ammonia nitrogen concentration.

7. Conclusion

To enhance the accuracy of water quality prediction, this paper designed a water quality prediction model using SVM algorithm, and compared it with the traditional neural network prediction method. The average concentration value of permanganate index and ammonia nitrogen concentration value of water source from January to June 2022 were observed, and the relative error was calculated. Finally, the feasibility conclusion was reached. In contrast to the traditional neural network algorithm, the use of SVM algorithm for water quality prediction has greatly improved the accuracy of water quality prediction, which has achieved the expected results. SVM algorithm in water quality prediction has a relatively wide range of application prospects, which can be used in the future to promote the use of SVM water quality prediction model on a large scale.

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] 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

[2] Huanhai Yang, Shue Liu. A prediction model of aquaculture water quality based on multiscale decomposition. Mathematical Biosciences and Engineering. (2021) 18(6): 7561-7579. https://doi.org/10.3934/mbe.2021374

[3] 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

[4] Haghiabi Amir Hamzeh, Ali Heidar Nasrolahi, Abbas Parsaie. Water quality prediction using machine learning methods. Water Quality Research Journal. (2018) 53(1): 3-13. https://doi.org/10.2166/wqrj.2018.025

[5] Ayaz Muhammad, Nasir-Uddin Khan. Forecasting of heavy metal contamination in coastal sea surface waters of the karachi harbour area by neural network approach. Nature Environment and Pollution Technology. (2019) 18(3): 719-733.

[6] Barzegar Rahim, Mohammad Taghi Aalami, Jan Adamowski. Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. Stochastic Environmental Research and Risk Assessment. (2020) 34(2): 415-433. https://doi.org/10.1007/s00477-020-01776-2

[7] Hassan Md Mehedi. Efficient prediction of water quality index (WQI) using machine learning algorithms. Human-Centric Intelligent Systems. (2021) 1(3-4): 86-97. https://doi.org/10.2991/hcis.k.211203.001

[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] Meshram Sarita Gajbhiye. Application of artificial neural networks, support vector machine and multiple model-ANN to sediment yield prediction. Water Resources Management. (2020) 34(15): 4561-4575. https://doi.org/10.1007/s11269-020-02672-8

[10] Lei Wang. Bathymetry retrieval from optical images with spatially distributed support vector machines. GIScience & remote sensing. (2019) 56(3): 323-337. https://doi.org/10.1080/15481603.2018.1538620

[11] Shamshirband Shahaboddin. Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines. Engineering Applications of Computational Fluid Mechanics. (2020) 14(1): 805-817. https://doi.org/10.1080/19942060.2020.1773932

[12] Seifi Akram, Hossien Riahi. Estimating daily reference evapotranspiration using hybrid gamma test-least square support vector machine, gamma test-ANN, and gamma test-ANFIS models in an arid area of Iran. Journal of Water and Climate Change. (2020) 11(1): 217-240. https://doi.org/10.2166/wcc.2018.003

[13] Zhi Wei. From hydrometeorology to river water quality: can a deep learning model predict dissolved oxygen at the continental scale? Environmental science & technology. (2021) 55(4): 2357-2368. https://doi.org/10.1021/acs.est.0c06783

[14] 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

[15] Changmin Sin. Now and Future of Water Quality Forecasting System of National Institute of Environmental Research." Water for future. (2019) 52(6): 22-26.

[16] Piasecki Adam, Jakub Jurasz. Forecasting daily water consumption: A case study in Torun, Poland. Periodica Polytechnica Civil Engineering. (2018) 62(3): 818-824. https://doi.org/10.3311/PPci.11930

[17] Ahmed Umair. Water quality monitoring: from conventional to emerging technologies. Water Supply. (2020) 20(1): 28-45. https://doi.org/10.2166/ws.2019.144

[18] Muharemi Fitore, Florin Leon. Machine learning approaches for anomaly detection of water quality on a real-world data set. Journal of Information and Telecommunication. (2019) 3(3): 294-307. https://doi.org/10.1080/24751839.2019.1565653

[19] Peterson Kyle T., Vasit Sagan, John J. Sloan. Deep learning-based water quality estimation and anomaly detection using Landsat-8/Sentinel-2 virtual constellation and cloud computing. GIScience & Remote Sensing. (2020) 57(4): 510-525. https://doi.org/10.1080/15481603.2020.1738061

[20] Gokulanathan S. A GSM based water quality monitoring system using Arduino. Shanlax International Journal of Arts, Science and Humanities. (2019) 6(4): 22-26. https://doi.org/10.34293/sijash.v6i4.341