

## Research Article

# Construction of Surface Water Pollution Prediction Model Based on Machine Learning

Yuanhong Che,<sup>1</sup> Zhangdong Wei ,<sup>1</sup> and Jing Wen <sup>2</sup>

<sup>1</sup>Miami College of Henan University, Kaifeng 475002, Henan, China

<sup>2</sup>Hunan Career Technical College Nonferrous Metal, Zhuzhou, Hunan 412006, China

Correspondence should be addressed to Zhangdong Wei; 10380012@vip.henu.edu.cn and Jing Wen; wenjing@cumt.edu.cn

Received 4 March 2022; Revised 24 March 2022; Accepted 7 April 2022; Published 28 April 2022

Academic Editor: Hangjun Che

Copyright © 2022 Yuanhong Che et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Existing prediction models have low prediction accuracy for surface water pollution with many influencing factors. Taking algal bloom prediction as the entry point of surface water pollution research, LASSO-LARS algorithm is adopted to select the main factors affecting algal bloom variables. At the same time, combined with BP neural network in machine learning, a surface water pollution prediction model based on machine learning is proposed. The results show that the proposed method can accurately predict the algal blooms by using the BP neural network algal bloom prediction model, where the LASSO-LARS algorithm is used to select the variables, such as water temperature, pH, transparency, conductivity, dissolved oxygen, ammonia nitrogen, and chlorophyll a, as model inputs, and the relative error of prediction is less than 5.2%. Thus, the proposed method has certain validity and practical application value.

## 1. Related Work

Surface water is an important water resource. However, with the acceleration of urbanization, the problem of water pollution is becoming more and more serious, which seriously affects the safety of human domestic water. In recent years, the problem of water quality safety has received attention from all walks of life. In order to continuously improve water quality and strengthen the safety of domestic water, various water treatment technologies and water management methods have emerged. Zhao et al. proposed an in situ loading method for the preparation of nanofiber composite aerogel by using aromatic nanofibers (ANFs) and prepared a high-performance 3D ANF/MnO<sub>2</sub> composite aerogel that can contain more than 90% Pb<sup>2+</sup> in water, which provides a promising insight for the treatment of Pb<sup>2+</sup> in water pollution [1]. Ma et al. combined chitosan base modified polymer (MCS) with Fe<sub>3</sub>O<sub>4</sub> and SiO<sub>2</sub> by silane coupling agent and successfully prepared a new type of recyclable covalently bonded magnetic flocculant (FS-MC), which can effectively remove emulsified oil in water [2]. Huang et al. used polydimethylsiloxane (PDMS) as the

dispersing phase and glutaraldehyde as the crosslinking agent, and Pickering emulsion stabilized by chitin nanofiber (ChNF) was adopted to prepare hydrophobic aerogel. Nonaqueous phase HOCs are selectively removed from water. It has broad application prospect in water treatment related to HOCs [3]. Saima prepared nitrogen-doped graphene oxide (NGO) nanocomposites (NGO/BQDS-TiO<sub>2</sub>) and bismuth oxide quantum dot (BQD)-doped TiO<sub>2</sub> composite photocatalyst, which are used to degrade 2, 4-dichlorophenol (2, 4-DCP) and other organic pollutants and stable dyes [4]. Sun et al. prepared a new type of surface molecularly imprinted polymer (SMIP), which can achieve selective separation and recovery of metal complex dyes in wastewater, providing a theoretical basis for the treatment of metal complex dyes in wastewater [5]. Saravanan et al. [6] discussed the treatment methods of various pollution sources such as heavy metals, dyes, pathogens, and organic matter in water pollution from physical, chemical, and biological aspects and believed that membrane separation and adsorption methods were more suitable for water pollution control. Doyle and Wu et al. prepared polyvinylidene fluoride (PVDF) composite ultrafiltration membrane (PGA

membrane) by blending modification of GO and adsorption and filtration of silver carbonate ( $\text{Ag}_2\text{CO}_3$ ) on the membrane surface, which can improve the anti-fouling performance and permeability of PVDF ultrafiltration membrane and effectively treat the slightly polluted surface water [7, 8]. As can be seen, the current research on water pollution is mainly focused on water pollution treatment, and there are fewer studies on prediction. Putri et al. used the Stella model to predict the pollution of the river, and the results show that this method can accurately predict the pollution of each section [9]. Wang et al. proposed a water pollution prediction method based on grey correlation, in which the main influencing factors of water pollution are found out through the grey correlation algorithm, and the water pollution is predicted by white differential equation. The results show that the prediction accuracy of this method is high and the error is small [10]. Wang et al. applied SVM algorithm to water pollution prediction. The results show that this method has good accuracy for small samples and nonlinear data [11]. In addition, the research on water pollution prediction is the key to water pollution treatment. Therefore, in order to better deal with water pollution, LASSO algorithm and BP neural network are adopted to propose a surface water pollution prediction model based on machine learning.

## 2. Basic Methods

**2.1. LASSO Algorithm.** LASSO algorithm is a variable selection method. Using the absolute value function of model coefficients as punishment, the regression coefficient is reduced. Moreover, variable selection and parameter estimation are achieved [12]. The main basis of LASSO's algorithm is the accuracy, interpretability, stability, and complexity of variable predictions to the model, and the coefficient of variables is punished. Assume that there are data  $(x^i, y_i)$ ,  $i = 1, 2, \dots, N$ ,  $X^i = (x_{i1}, \dots, x_{ip})^T$  and  $y_i$  is the independent variable and response variable corresponding to the observed value  $i$ . The linear regression model can be expressed as follows:

$$y_i = a_i + \sum_{j=1}^p \beta_j x_{ij} + \varepsilon_i, \quad \varepsilon_i \in N(0, \sigma^2). \quad (1)$$

Generally, in regression structures, conditions of  $y_i$  and  $x^i$  are independent [13]. Therefore,  $x_{ij}$  is assumed to be standardized; then, there are

$$\begin{aligned} \frac{1}{N} \sum_j x_{ij} &= 0, \\ \frac{1}{N} \sum_j x_{ij}^2 &= 0. \end{aligned} \quad (2)$$

When  $t \geq 1$ , LASSO's estimate is

$$(\hat{\alpha}, \hat{\beta}) = \arg \begin{cases} \min_{\beta} \sum_i \left( y_i - a_i - \sum_i \beta_j x_{ij} \right)^2, \\ \text{s.t.} \quad \sum_j |\beta_j| \leq t. \end{cases} \quad (3)$$

Although LASSO algorithm has some advantages in theory compared with traditional variable selection methods such as partial least squares method and principal component analysis method, it has the problem of large computation [14]. To solve this problem, Efron reselects variables on each regression residual to reduce the residual, and LARS algorithm is proposed to reduce the computation of LASSO algorithm [15]. Therefore, the LASSO-LARS algorithm is used to select influencing factor variables of surface water pollution.

**2.2. BP Network.** BP neural network is a typical neural network, which carries out network training through error backpropagation. Its basic structure is shown in Figure 1, consisting of three layers: input layer, output layer, and hidden layer [16, 17].

Assume that the input vector is  $x_i = (x_1, x_2, \dots, x_M)^T$ , the output vector is  $O_j = (y_1, y_2, \dots, y_J)^T$ , and the connection weights between input layer and hidden layer and between hidden layer and output layer are  $\omega_{mi}$  ( $i = 1, 2, \dots, I$ ) and  $\omega_{ij}$  ( $j = 1, 2, \dots, J$ ), respectively. Therefore, for hidden layer, there are

$$y_i = f(\text{net}_i), \quad i = 1, 2, \dots, I, \quad (4)$$

$$\text{net}_i = \sum_{l=1}^M x_l \omega_{li}, \quad i = 1, 2, \dots, I. \quad (5)$$

For output layer, there are

$$o_k = f(\text{net}_k), \quad k = 1, 2, \dots, J, \quad (6)$$

$$\text{net}_k = \sum_{i=1}^I y_i \omega_{ik}, \quad k = 1, 2, \dots, J. \quad (7)$$

In formulas (4) and (6),  $f(x)$  is the transfer function, which usually is sigmoid function and expressed as [18]

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (8)$$

The derivative function is

$$f'(x) = f(x)(1 - f(x)). \quad (9)$$

BP neural network adopts the gradient descent method to adjust weights, and the adjustment formula is [19]

$$\begin{cases} \Delta \omega_{ik} = \eta \delta_{ik}^{\omega} y_i, \\ \Delta \omega_{li} = \eta \delta_{li}^{\omega} x_l. \end{cases} \quad (10)$$

Based on the above analysis, it can be seen that the LASSO-LARS algorithm can ideally select the factor variables affecting surface water pollution, which determines the main influencing factors for accurate prediction of surface water pollution. However, the BP neural network has the characteristics of fast search speed and strong local optimization ability [20]. Therefore, variables selected by LASSO-LARS algorithm are taken as the input of prediction model, and a surface

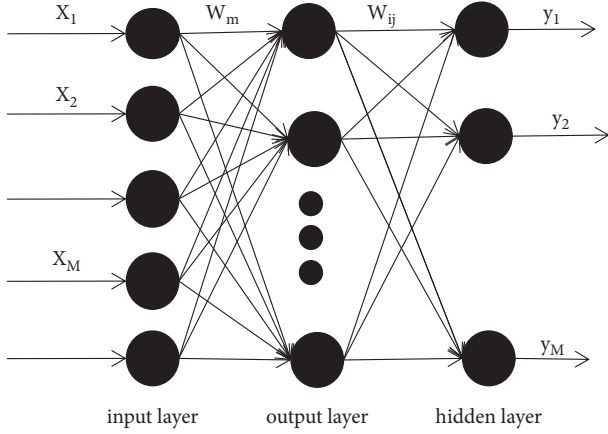


FIGURE 1: Basic structure of BP neural network.

water pollution prediction model based on BP neural network is constructed.

### 3. Surface Water Pollution Prediction Model Based on Machine Learning

*3.1. Selection of Influencing Factor Variables of Surface Water Pollution Based on LASSO-LARS.* Algal bloom is one of the main causes of surface water pollution. Algal bloom pollution is taken as a representative to explore surface water pollution prediction based on machine learning. There are many factors affecting algae blooms, including 234 indexes such as water temperature, dissolved oxygen, and metal ions [21]. Among them, 16 indexes such as water temperature, transparency, and chlorophyll a, as shown in Table 1, are the main factors affecting algae blooms [22].

As can be seen, the above 16 indicators are selected as available variables, and logarithmic processing is performed on them, which are represented as  $\ln Y, \ln X_1, \ln X_2, \dots, \ln X_{16}$ . Moreover, the established linear model is shown as follows:

$$\ln Y = a_0 + a_1 \ln X_1 + a_2 \ln X_2 + \dots + a_{16} \ln X_{16} + \varepsilon, \quad (11)$$

where  $a_0$  is the constant term;  $a_1, \dots, a_{16}$  represent the variable coefficient; and  $\varepsilon$  is the random disturbance term.

LASSO-LARS algorithm is used to calculate the above influencing factors. According to the Akaike information criterion, the smaller the statistical value is, the easier the optimal solution is [23]. Figure 2 shows the values of the variables for which the current model is optimal.  $X_1, X_3, X_5, X_7, X_9, X_{12}$ , and  $X_{14}$  have a great influence on algae blooms, namely, water temperature, pH, transparency, conductivity, dissolved oxygen, ammonia nitrogen, and chlorophyll a. Therefore, the above 7 variables are selected as the main factor variables of algae blooms in surface water.

*3.2. Surface Water Pollution Prediction Model Based on BP.* Based on the above factors and BP network selected by LASSO-LARS, the construction process of algal outbreak prediction model is designed as follows.

TABLE 1: Selection and definition of influencing factor variables of algae blooms.

Variable name	Variable definition
X1	T (°C)
X2	Fe (mg/L)
X3	pH
X4	Mn (mg/L)
X5	SD (m)
X6	Cu (mg/L)
X7	Sp Cond ( $\mu\text{s}/\text{cm}$ )
X8	Zn (mg/L)
X9	DO (mg/L)
X10	TN (mg/L)
X11	TP (mg/L)
X12	$\text{NH}_4^+$ (mg/L)
X13	Cd (mg/L)
X14	Chlorophyll a (mg/L)
X15	Pb (mg/L)
X16	Hg (mg/L)

*3.2.1. Determination of Sample.* The collected surface water quality data can be divided into training sample set and test sample set according to certain rules. Among them, the training sample set is used for BP neural network learning and training, so as to obtain an optimal BP neural network prediction model. The test sample set is used to verify whether the prediction model constructed by the test can achieve the ideal prediction effect for unfamiliar data.

*3.2.2. Determination of Network Structure.* In general, the standard three-layer BP neural network can achieve relatively accurate prediction [24]. Therefore, three-layer network structure is adopted to construct the prediction model. According to the 7 factors influencing algae blooms selected by LASSO-LARS, the number of neurons in the input layer of BP neural network prediction model can be determined to be 7. Since the algal Q index is the main indicator to measure the extent of algal blooms, the algal Q index is used as the model output [25], and the number of neurons in the output layer is 1. The number of neurons  $n$  in the hidden layer can be determined by formula (12). After repeated training, it can be seen that  $n=3$ .

$$n = \sqrt{P + Q} + \alpha, \quad (12)$$

where  $n$  is the number of neurons in the hidden layer,  $P$  and  $Q$  represent the number of neurons in the input and output layer, and  $\alpha \in [1, 10]$ .

*3.2.3. Data Preprocessing.* Due to the different dimensions of different data types in the sample dataset, the input requirements of BP neural network prediction model are not met, and there are problems such as missing values, which affect the final prediction effect of the model. Therefore, before the sample is input into the model, the data need to be preprocessed by normalization, standardization, missing value filling, and so on.

TABLE 2: Variable selection based on LASSO-LARS.

Coefficient	1	2	3	4	5	6	7	8	9	10
$\alpha_1$	0	0.311	0.317	0.319	0.682	0.557	0.56	0.542	0.542	0.53
$\alpha_2$	0	0	0	0	0	0.11	0	0	0	0
$\alpha_3$	0	0	0	0	0	0	0.14	0.126	0.128	0.116
$\alpha_4$	0	0	0	0	0	0	0	0.026	0	0
$\alpha_5$	0	0	0	0	0	0	0	0.006	0.004	0.002
$\alpha_6$	0	0	0	0	0	0	0	0	0	0
$\alpha_7$	0	0	0	0	0	0	0	0.008	0.009	0.005
$\alpha_8$	0	0	0	0	0	0	0	0	0	0
$\alpha_9$	0	0	0	0	0	0.132	0.137	0.159	0.177	0.196
$\alpha_{10}$	0	0	0	0	0	0	0.121	0	0	0
$\alpha_{11}$	0	0	0	0	0.04	0	0	0	0	0
$\alpha_{12}$	0	0	0	0	0	0	0	0.195	0.185	0.116
$\alpha_{13}$	0	0	0	0	0	0.014	0	0	0	0
$\alpha_{14}$	0	0.206	0.329	0.322	0.287	0.297	0.18	0.196	0.208	0.291
$\alpha_{15}$	0	0	0	0.102	0	0	0	0	0	0
$\alpha_{16}$	0	0	0	0	0	0	0.003	0	0	0
AIC	2.581	3.476	0.964	0.074	0.056	0.034	0.031	0.026	0.022	0.012

3.2.4. *Model Construction and Prediction.* The BP neural network prediction model can be constructed by calling Newff function in MATLAB software, and then the best prediction model retained through repeated training can be used to predict algal blooms. The specific prediction steps are as follows. According to the 7 main factors affecting the algae outbreak, the prediction variables are established by using the obtained data, and then the prediction variables are input into prediction model to obtain the output of variables, namely, the prediction result.

## 4. Simulation Experiment

4.1. *Experimental Environment.* MATLAB Newff function is called to construct the surface water pollution prediction model, which is run on Windows 7 operating system, and Premnmx function is called to normalize the input data.

4.2. *Data Sources and Preprocessing.* The experimental dataset is collected from the average array of weekly water quality data of a certain water sample from March to November in Daqing city, Heilongjiang Province, from 2003 to 2020 (a total of 648 groups), and the specific collection work is divided into two parts: on-site monitoring and laboratory testing analysis. Water temperature must be measured in the field, and chlorophyll a must be taken back to the laboratory for testing. In order to make the input data meet the input requirements of model, centralization is performed on the input model data by using formulas (13)–(17); the input model data are normalized by using formula (16); and the output model data are reversely normalized by using formula (17) [26–28].

$$\sum_{i=1}^n y_i = 0, \quad (13)$$

$$\sum_{i=1}^n x_{ij} = 0, \quad (14)$$

$$\sum_{i=1}^n x_{ij}^2 = 1, \quad (15)$$

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad i = 1, 2, \dots, n, \quad (16)$$

$$x_i = \bar{x}_i (x_{\max} - x_{\min}) + x_{\min}. \quad (17)$$

In formulas (16) and (17),  $x_i$  represents the input or output data and  $x_{\min}$  and  $x_{\max}$  are the minimum and maximum values.

576 groups of preprocessed experimental data from 2003 to 2018 are selected as training datasets, and the remaining 72 groups of data are selected as test datasets.

4.3. *Evaluation Indicators.* Relative error  $\delta$  is selected as an indicator to evaluate the prediction performance of model, and its calculation method is shown in the following formula:

$$\delta = \frac{\Delta}{L * 100\%}, \quad (18)$$

where  $\Delta$  is the absolute error, which can be calculated by formula (19), and  $L$  is the real value.

$$\Delta = X - L, \quad (19)$$

where  $X$  is the measured value and  $L$  is the real value.

### 4.4. Experimental Results

4.4.1. *Variable Selection Result.* To verify the effect of LASSO-LARS algorithm on variable selection, nine groups of water quality data in the first week of each month from March to November 2020 are used to illustrate.

(1) *Water Temperature.* Figure 2 shows the distribution characteristics of water temperature in the first week of each

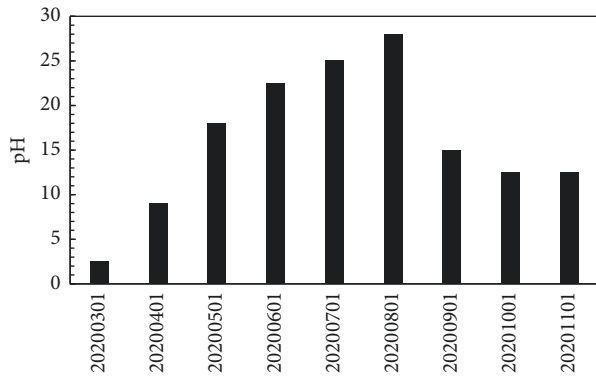


FIGURE 2: Distribution characteristics of water temperature.

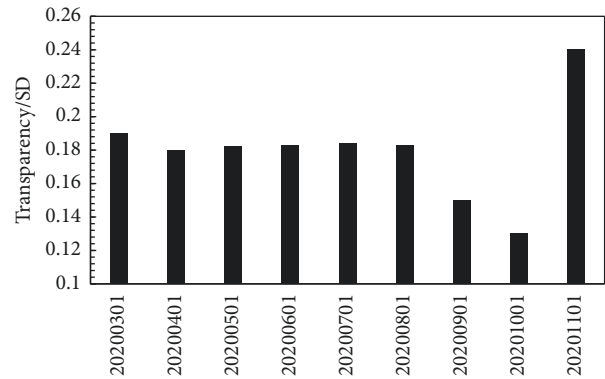


FIGURE 4: Distribution characteristics of transparency.

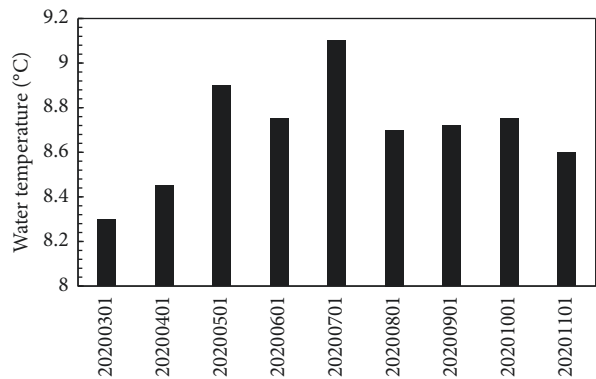


FIGURE 3: Distribution characteristics of pH.

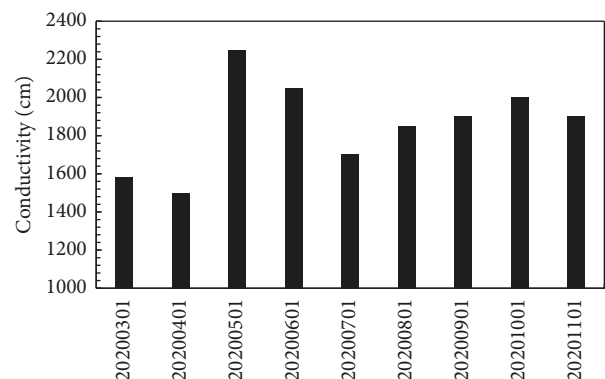


FIGURE 5: Distribution characteristics of conductivity.

month from March to November 2020. As can be seen, the water temperature changes significantly with the seasons. In 2020, the water temperature in this region is below 10°C in March, April, and November and above 10°C from May to September, with an average temperature of 15.42°C. According to the literature, cyanobacteria have certain growth advantages in water temperature of 40°C; Chlorophyta has certain growth advantages in water temperature of 30°C; and diatoms have certain growth advantages in water temperature of 20°C. Overall, the species and quantity of algae are closely related to the water temperature, and the water temperature varies significantly with the season, leading to the seasonal variation of algae species. Thus, there is a strong correlation between water temperature and algal blooms.

(2) *pH*. Figure 3 shows the pH distribution characteristics of water body in the first week of each month from March to November 2020. As can be seen, the pH of the water varies between 8.30 and 9.18, and it is slightly alkaline on the whole.

(3) *Transparency*. Figure 4 shows the transparency distribution characteristics of water body in the first week of each month from March to November 2020. As can be seen, the transparency of this water body is mostly concentrated at 0.18 m, which is generally low. The lowest water transparency is 0.13 m in October, and the highest is 0.24 m in November.

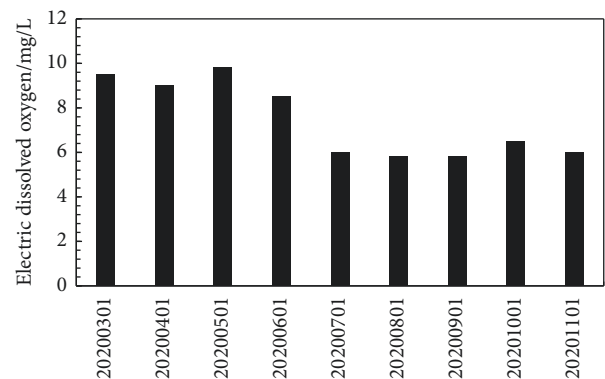


FIGURE 6: Distribution characteristics of dissolved oxygen.

(4) *Conductivity*. Figure 5 shows the conductivity of water body during the study period. In this water environment, the conductivity is generally high, reaching more than 1600  $\mu\text{s/cm}$ . The reason is that the selected water is close to the sewage inlet and contains a large amount of industrial wastewater and domestic sewage.

(5) *Dissolved Oxygen*. Dissolved oxygen is an important index of self-clean capacity of the water body, and its concentration reflects the growth of algae. Figure 6 shows the dissolved oxygen concentration of water body in the study area in different periods. As can be seen, the dissolved oxygen concentration of water body presents a wavy

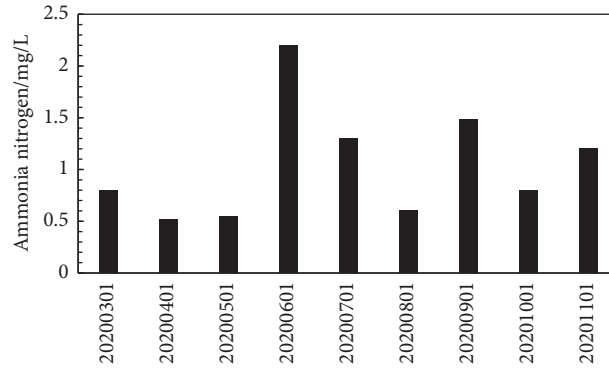


FIGURE 7: Distribution characteristics of ammonia nitrogen.

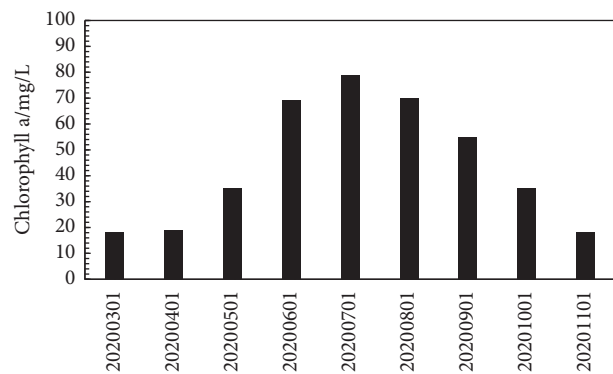


FIGURE 8: Distribution characteristics of chlorophyll a.

TABLE 3: Real values of test sample data.

Time	Water temperature (°C)	pH	Transparency (cm)	Conductivity ( $\mu\text{s}/\text{cm}$ )	Dissolved oxygen (mg/L)	Ammonia nitrogen (mg/L)	Chlorophyll a (mg/L)	Algal Q index
20200301	2.21	8.31	19	1571	9.51	0.74	17.25	2.54
20200401	9.05	8.47	18	1482	8.92	0.59	18.04	1.34
20200501	17.19	8.87	18.2	2260	9.78	0.63	35.91	1.31
20200601	21.15	8.76	18.35	2055	8.47	2.04	68.58	0.71
20200701	25.23	9.15	18.51	1706	6.02	1.21	78.31	0.57
20200801	27.72	8.71	18.42	1831	5.47	0.59	69.83	0.60
20200901	15.11	8.73	15.24	1936	5.52	1.47	55.05	0.48
20201001	12.46	8.74	13.12	2005	6.32	0.79	34.84	1.85
20201101	8.30	8.61	24.33	1948	5.91	1.12	15.21	2.80

TABLE 4: Comparison between predicted value and real value of the model.

Time	Real value (mg/L)	Predicted value (mg/L)	Absolute error (mg/L)	Absolute error (%)
20200301	2.54	2.64	0.10	4.18
20170401	1.34	1.28	-0.06	4.92
20200501	1.31	1.37	0.06	4.81
20200601	0.71	0.73	0.02	3.68
20200701	0.57	0.54	-0.03	4.73
20200801	0.60	0.63	0.03	4.85
20200901	0.48	0.50	0.02	4.42
20201001	1.85	1.93	0.08	4.71
20201101	2.80	2.65	-0.15	5.19

distribution with seasonal changes, with a variation range of [5.47, 9.78] (unit: mg/L). According to the analysis, when the water surface is wide and agitated, the concentration of dissolved oxygen is high; when the water surface is small and without fluctuation, the concentration of dissolved oxygen is low, and too low dissolved oxygen will lead to the death of a large number of algae and deteriorate the water quality.

(6) *Ammonia Nitrogen*. Figure 7 shows the ammonia nitrogen concentration in water body during the study. As can be seen, ammonia nitrogen concentration in this water body reaches its lowest and highest values in April and June, with obvious seasonal variations.

(7) *Chlorophyll a*. Chlorophyll a concentration is positively correlated with water pollution degree. Figure 8 shows the concentration of chlorophyll a in water body during the study period. It can be seen from the figure that during the study period, the pollution level in the water is high and algae growth is active.

In conclusion, the 7 influencing factor variables selected by LASSO-LARS algorithm are closely related to algae blooms, which indicates that LASSO-LARS algorithm is effective.

**4.4.2. Prediction Results of the Model.** There are 576 sets of data from the training set adopted to train BP surface water pollution prediction model, and the best prediction model with the minimum error is saved. 72 groups of test data are input into the best prediction model and compared with the real value, and thus the prediction effect of model can be obtained. Due to the large amount of data, only data of the first week of each month from March to November 2020 are presented, as shown in Table 3, and the evaluation indicator values predicted by the model are shown in Table 4. As can be seen from Table 3, the water environment from March to May in 2020 is generally good, with mild pollution and certain tolerance. From June to October, the water quality changed from moderate pollution to severe pollution and eventually returned to moderate pollution. Water body is slightly polluted in November.

As can be seen from Table 4, there is a certain error between predicted value and real value. The absolute error is more than 0.1 in all months except March and November 2020. The relative error is between 3.6% and 5.2%, which is within the allowable error range. Therefore, the proposed prediction model can predict surface water pollution well with high prediction accuracy and good universality.

## 5. Conclusion

To sum up, the proposed surface water pollution prediction model based on machine learning can accurately predict the algae blooms. Here the BP prediction model is established, and the LASSO-LARS algorithm is utilized to select water temperature, pH, transparency, conductivity, dissolved oxygen, ammonia nitrogen, and chlorophyll a variables as the model input. Furthermore, the relative prediction error

is less than 5.2%, which means that the prediction accuracy is higher. All in all, it provides a theoretical basis for the prediction of surface water pollution. However, due to the limitations of conditions, there are still some problems to be improved. When using BP neural network to construct the surface water pollution prediction model, there are many methods to improve the accuracy of BP network model, for example, optimizing the network parameters by genetic algorithm can improve the prediction accuracy of model to a certain extent. Therefore, to better realize surface water pollution prediction, the next research is to use the above algorithm to improve BP neural network or explore a new prediction model.

## Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This study was sponsored in part by the Science and Education Joint Project of Hunan Natural Science Foundation (2020JJ7058).

## References

- [1] G. Zhao, H. Zhao, L. Shi, B. Cheng, X. Xu, and X. Zhuang, "In situ loading MnO<sub>2</sub> onto 3D Aramid nanofiber aerogel as High-Performance lead adsorbent," *Journal of Colloid and Interface Science*, vol. 600, pp. 403–411, 2021.
- [2] J. Ma, X. Fu, W. Xia et al., "Removal of emulsified oil from water by using recyclable chitosan based covalently bonded composite magnetic flocculant: performance and mechanism," *Journal of Hazardous Materials*, vol. 419, pp. 126529–126539, 2021.
- [3] Y. Huang, H. Sun, and H. Liu, "Fabrication of chitin nanofiber-PDMS composite aerogels from Pickering emulsion templates with potential application in hydrophobic organic contaminant removal," *Journal of Hazardous Materials*, vol. 419, pp. 126475–126485, 2021.
- [4] N. Saima, "Competitive role of nitrogen functionalities of N doped GO and sensitizing effect of Bi<sub>2</sub>O<sub>3</sub> QDs on TiO<sub>2</sub> for water remediation," *Journal of Environmental Sciences*, vol. 108, pp. 107–119, 2021.
- [5] Y. Sun, Y. Gu, and Q. Zha, "A novel surface imprinted resin for the selective removal of metal-complexed dyes from aqueous solution in batch experiments: ACB GGN as a representative contaminant," *Chemosphere*, vol. 280, pp. 130611–130617, 2021.
- [6] A. Saravanan, P. Senthil Kumar, S. Jeevanantham et al., "Effective water/wastewater treatment methodologies for toxic pollutants removal: processes and applications towards sustainable development," *Chemosphere*, vol. 280, pp. 130595–130598, 2021.
- [7] S. Doyle, E. Meade, J. Gao et al., "A rapid antimicrobial photodynamic water treatment strategy utilizing a xanthene

- dye with subsequent removal by Goethite Nanoparticles,” *Chemosphere*, vol. 280, pp. 130764–130768, 2021.
- [8] L. Wu, Y. Liu, J. Hu, X. Feng, C. Ma, and C. Wen, “Preparation of polyvinylidene fluoride composite ultrafiltration membrane for micro-polluted surface water treatment,” *Chemosphere*, vol. 284, pp. 131294–131297, 2021.
- [9] F. A. Putri and N. Karnaningroem, “Prediction of water pollution in Kali Surabaya river segment Karangpilang-Ngagel using stella model,” *IOP Conference Series: Earth and Environmental Science*, vol. 259, no. 1, Article ID 012010, 2019.
- [10] R. Wang, F. X. Yang, and G. M. Qu, “Prediction model of agricultural non-point source water pollution based on grey correlation method,” *Nature Environment and Pollution Technology*, vol. 20, no. 2, pp. 633–641, 2021.
- [11] Y. Wang, S. Xue, and J. Ding, “Research on water pollution prediction of township enterprises based on support vector regression machine,” *E3S Web of Conferences*, vol. 228, Article ID 02014, 2021.
- [12] P. Kittisuwan, “Analytical and simple form of shrinkage functions for non-convex penalty functions in fused lasso algorithm,” *The International Journal on Artificial Intelligence Tools*, vol. 29, no. 6, pp. 15–18, 2020.
- [13] J. Lv and M. Pawlak, “Additive modeling and prediction of transient stability boundary in large-scale power systems using the Group Lasso algorithm,” *International Journal of Electrical Power & Energy Systems*, vol. 113, no. C, pp. 963–970, 2019.
- [14] S. Datta, V. A. Dev, and M. R. Eden, “Using correlation based adaptive LASSO algorithm to develop QSPR of antitumour agents for DNA–drug binding prediction,” *Computers & Chemical Engineering*, vol. 122, pp. 28–264, 2019.
- [15] V. Bochkarev, V. Tyurin, A. Savinkov, and B. Gizatullin, “Application of the LASSO algorithm for fitting the multi-exponential data of the NMR relaxometry,” *Journal of Physics: Conference Series*, vol. 1141, no. 1, pp. 012148–012158, 2018.
- [16] B. Ke, H. Nguyen, X.-N. Bui, H.-B. Bui, and T. Nguyen-Thoi, “Prediction of the sorption efficiency of heavy metal onto biochar using a robust combination of fuzzy C-means clustering and back-propagation neural network,” *Journal of Environmental Management*, vol. 293, pp. 112808–112818, 2021.
- [17] S. Jegan, T. B. Pushpa, and R. Gokulan, “Artificial neural network modelling for biodecolorization of Basic Violet 03 from aqueous solution by biochar derived from agro-bio waste of groundnut hull: kinetics and thermodynamics,” *Chemosphere*, vol. 276, pp. 130191–130196, 2021.
- [18] H. Wang, Q. M. J. Wu, D. Wang, J. Xin, Y. Yang, and K. Yu, “Echo state network with a global reversible autoencoder for time series classification,” *Information Sciences*, vol. 570, pp. 744–768, 2021.
- [19] M. Horvat, K. Pastor, and P. Kristian, “Multivariate analysis of water quality parameters in Lake Palic, Serbia,” *Environmental Monitoring and Assessment*, vol. 193, no. 7, pp. 410–414, 2021.
- [20] M. Mario and P. Daniela, “Multivariate analysis of water quality data for drinking water supply systems,” *Water*, vol. 13, no. 13, p. 1766, 2021.
- [21] Z. S. Peng, H. C. Ji, W. C. Pei, B. Y. Liu, and G. Song, “Constitutive relationship of TC4 titanium alloy based on back propagating (BP) neural network (NN),” *Metalurgija*, vol. 60, no. 3-4, pp. 277–280, 2021.
- [22] H. Cui, Q. Zhang, J. Zhang, Z. Wu, and W. Wu, “Classification of grain storage inventory modes based on temperature contour map of grain bulk using back propagation neural network,” *Agriculture*, vol. 11, no. 5, pp. 451–457, 2021.
- [23] Y. Jin, S. Yu, D. H. Kim, E. J. Yun, and K. H. Kim, “Characterization of neogaroooligosaccharide hydrolase BpGH117 from a human gut bacterium *Bacteroides plebeius*,” *Marine Drugs*, vol. 19, no. 5, pp. 271–275, 2021.
- [24] F. Ruan, X. Ding, H. Li, Y. Wang, K. Ye, and H. Kan, “Back propagation neural network model for medical expenses in patients with breast cancer,” *Mathematical Biosciences and Engineering*, vol. 18, no. 4, pp. 3690–3698, 2021.
- [25] H. He, “Learnability and robustness of shallow neural networks learned by a performance-driven BP and a variant of PSO for edge decision-making,” *Neural Computing & Applications*, vol. 33, no. 20, pp. 1–22, 2021.
- [26] H. Liu, J. Liu, Y. Wang, Y. Xia, and Z. Guo, “Identification of grouting compactness in bridge bellows based on the BP neural network,” *Structures*, vol. 32, pp. 817–826, 2021.
- [27] F. Chun, N. Zhao, G. Yin et al., “Artificial neural networks combined multi-wavelength transmission spectrum feature extraction for sensitive identification of waterborne bacteria,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 251, pp. 119423–119433, 2021.
- [28] F. Yu, “An active construction dynamic schedule management model: using the fuzzy earned value management and BP neural network,” *KSCE Journal of Civil Engineering*, vol. 68, pp. 1–15, 2021.