Research on application of ReliefF and improved RVM in water quality grade evaluation

Yu-Lin Gong*, Ming-Jia Hu, Hui-Fang Yang and Bo Han
School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun, Jilin, China
*Corresponding author. E-mail: yulingong000@163.com

ABSTRACT

ReliefF algorithm was used to analyze the weight of each water quality evaluation factor, and then based on the Relevance Vector Machine (RVM), Particle Swarm Optimization (PSO) was used to optimize the kernel width factor and hyperparameters of RVM to build a water quality evaluation model, and the experimental results of RVM, PSO-RVM, ReliefF-RVM and PSO-ReliefF-RVM were compared. The results show that ReliefF algorithm, combined with threshold value, selects 5 evaluation factors with significant weight from eight evaluation factors, which reduces the amount of data used in the model, CSI index is used to calculate the separability of each evaluation factor combination. The results show that the overall separability of the combination is best when the evaluation factor with significant weight is reserved. When different water quality evaluation factors were included, the evaluation accuracy of PSO-ReliefF-RVM model reached 95.74%, 14.23% higher than that of RVM model, which verified the effectiveness of PSO algorithm and ReliefF algorithm, and had a higher guiding significance for the study of water quality grade evaluation. It has good practical application value.

Key words: particle swarm optimization, relevance vector machine, ReliefF algorithm, water quality grade evaluation

HIGHLIGHT

• By comparing the prediction results of PSO-BP neural network and AdaBoost-BP neural network, it is found that the AdaBoost-BP evaluation model has higher accuracy, better stability and better application value.

In industrialized highly developed today, the factory every day to produce a large number of industrial wastewater, but due to the small and medium-sized factories generally weak environmental awareness, a large number of untreated industrial wastewater directly discharge into the rivers, the eutrophication of water bodies, wetland reduce, aquatic organisms and a series of problems, water ecological environment worsening, and the governance of water pollution are imperative. Water quality evaluation is one of the main contents of water environment quality research, which provides scientific basis for rational development and utilization of water environment and comprehensive prevention and control of water pollution (Yongcai et al. 2021).

For the evaluation of water quality grade, mathematical statistical methods have been developed from the original biological evaluation method, expert evaluation method (Jingsheng 2000), single factor evaluation method (Lei & Xiaojun 2006), to the present fuzzy comprehensive evaluation method (Zhijun 2020), fuzzy clustering analysis method (Xiaoyan et al. 2021), and principal component analysis method (Hong et al. 2017). The calculation method of single factor evaluation method is simple, but only the water quality evaluation method with single information is adopted. For the current complex and changeable water quality situation, the evaluation results often can’t fully reflect the actual comprehensive water quality category. The fuzzy comprehensive evaluation method is limited by the uncertainty of index weight, fuzzy operator, maximum membership principle and other aspects in the evaluation process, which limits the universality of the evaluation mode of this method in the field of water quality evaluation, and also affects the reliability of the evaluation results (Jing 2014; Yan et al. 2021). Fuzzy clustering analysis has the characteristics of intuitiveness and simple conclusion form, but it is difficult to obtain the clustering conclusion when dealing with large sample size data. Principal component analysis is more accurate for the analysis of a single source of pollution, but it cannot fully reflect the overall situation of water quality (Yang et al. 2021). To make up for the deficiency of statistical models, machine learning and other methods have been introduced into water quality monitoring, such as BP neural network (Guangyue 2020), support vector machine (Yan et al. 2021), random forest

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).
In the actual evaluation process of water quality grade, there will be a variety of water quality evaluation factors involved. The weight of different water quality evaluation factors is different, and their contribution to the evaluation of water quality grade is also different, and some evaluation factors even play a negative role in the evaluation process. At present, in the process of processing multiple evaluation factors, common methods include weight assignment method, dimension reduction method and feature reduction. Specific methods include Delphi method (Yuanyuan et al. 2013), analytic hierarchy process (Xiaojuan 2021), entropy weight method (Henghua et al. 2021), principal component analysis method (Yongjun et al. 2021), decision tree method (Guozhong et al. 2019), etc. By calculating the weight of each evaluation factor, the weighting method emphasizes the importance of the severely polluted water quality index in all indicators and increases its influence on the final score at the same time (Weiguo et al. 2019), but does not take into account the negative effects caused by some evaluation factors. Dimension reduction methods, such as principal component analysis, reduce the original dimension of multiple variables into a statistical analysis method of a small number of comprehensive variables, so as to retain all the original information to the greatest extent and reduce the complexity of calculation (Tengfei et al. 2018). However, such methods ignore the interpretation of independent variables to the dependent variable system. Feature selection methods such as decision tree can automatically detect and estimate the interaction effects among numerous independent variables, and are not affected by multicollinearity, and can better deal with extreme values and missing values (Lipin et al. 2020). However, decision tree has poor generalization ability of data and is prone to over-fitting.

ReliefF algorithm was used to calculate the weight of eight evaluation factors, and the evaluation factor with high weight was reserved as the final evaluation factor. The correlation vector machine was used to build the water quality evaluation model. CSI is used to calculate the separability of the combinations of different evaluation factors. The results show that the combination has the best separability when the five weighting factors are retained. Since the initial value of RVM was difficult to determine, the PSO algorithm was used to optimize the core width factor and hyperparameters of the RVM model, and the water quality evaluation models of RVM, PSO-RVM, ReliefF-RVM and PSO-ReliefF-RVM were established respectively. The experimental results show that PSO-ReliefF-RVM model can evaluate water quality more accurately than RVM model and has good practical application value.

1. WATER QUALITY DATA SOURCES

China has a large amount of water resources. Affected by natural environment and human factors in different regions, the water quality is also different. Therefore, in order to ensure scientific rigor of experimental data, this paper obtained water quality data from the National Surface Water Quality Automatic Monitoring Real-time System (Ministry of Ecology & Environment). By July 2021, the monitoring system had monitored 3,544 surface water national examination sections (point sites), including 3,205 river sections and 341 lake and reservoir points. A total of 1,824 rivers and 210 lakes and reservoirs were involved in seven river basins of China, effectively ensuring the diversity and effectiveness of experimental data.

According to the Surface Water Environmental Quality Standard (GB 3838-2002 2002), the national surface water quality automatic monitoring real-time system divides the water quality into class I to class V (poor V) according to the standard. Water quality evaluation involves many indicators, including eight evaluation indicators, respectively: PH value, ammonia nitrogen, dissolved oxygen, conductivity, turbidity, permanganate index, total phosphorus and total nitrogen. Among them, ammonia nitrogen is the product of aerobic decomposition of organic matter, which can lead to water eutrophication and is the index of water eutrophication. Dissolved oxygen refers to dissolved oxygen in water, to reflect the index of water self-purification capacity. Potassium permanganate index is similar to chemical oxygen demand and is also a comprehensive index reflecting organic pollution. Total phosphorus and total nitrogen are the concentrations of phosphorus and nitrogen in water, and are indicators of eutrophication in water. Turbidity of water refers to the turbidity situation caused by a large number of visible suspended substances in water samples. Conductivity shows how well water conducts electricity and how much salt it contains. Table 1 shows the corresponding water quality grade of each index value:

With the aid of national surface water quality automatic monitoring real-time system, this article from the selection of water quality automatic monitoring system in real-time to the multiple basin surface water quality data, a total of 3,760 sets, according to the experiment requirement, data is divided into training set and testing set data, and keep the same in four different evaluation network training set and testing set of data.
2. RELEVANCE VECTOR MACHINE

2.1. Basic theory

Correlation vector machine is a machine learning model based on sparse bayesian framework theory (Tipping 2001). Compared with Support Vector Machines (SVM), it has the following advantages: (1) correlation vectors are sparse; (2) training time can be saved by setting the nuclear parameters; (3) the flexibility of selecting kernel functions is increased so that kernel functions do not have to meet mercer conditions (Nahman & Salamon 2017).

Suppose that the input and output vectors of RVM are respectively \( \{x_u\}_{u=1}^N \) and \( \{t_u\}_{u=1}^N \), the relationship between the input vector and the target value shows as follow:

\[
I_u = \sum_{i=1}^N w_i K(x, x_i) + w_0 + \epsilon_u
\]

(1)

where \( \omega = (\omega_1, \omega_2, \cdots, \omega_N)^T \) and \( K(x, x_i) \) are kernel functions, \( \omega \) is corresponding weights, \( \epsilon_u \) is noise with mean zero and variance \( \sigma^2 \), and \( w_0 \) is constant.

Assuming that \( \{t_u\}_{u=1}^N \) is an independent random variable, the conditional probability of the target value \( t \) is:

\[
p(t|\omega, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{||t - \Phi\omega||^2}{2\sigma^2}\right)
\]

(2)

where \( t = (t_1, t_2, \cdots, t_N)^T \), \( \Phi \) is the kernel function matrix of \( N \times (N + 1) \).

To avoid overfitting of \( \omega \) and \( \sigma^2 \), constraint parameters of zero-mean gaussian prior probability distribution are often used:

\[
p(\omega|\alpha) = \prod_{i=0}^N N(\omega_i|0, \alpha_i^{-1})
\]

(3)

where \( \alpha \) is a hyperparameter vector of \( N + 1 \).

Based on Bayes formula, the weight of the posterior distribution is described as follows:

\[
p(\omega|t, \alpha, \sigma^2) = (2\pi)^{-N(N+1)/2} \prod_{i=0}^N \left| \sum_{\omega}^{-1/2} \exp\left\{ -\frac{1}{2}(\omega - \mu)^T (\sigma^{-2}\Phi^T\Phi + A)^{-1}(\omega - \mu) \right\} \right|
\]

(4)

Among them, \( \mu = \sigma^{-2} \sum \phi^T t \), \( \sum = (\sigma^{-2}\Phi^T\Phi + A)^{-1} \), \( A = \text{diag}(\alpha_0, \alpha_1, \cdots, \alpha_N) \).
To establish a unified hyperparameter, \( p(t; \alpha, \sigma^2) \) is defined as:

\[
p(t; \alpha, \sigma^2) = \int p(t; \omega, \sigma^2) p(\omega; \alpha) = \frac{1}{\sqrt{2\pi}} (\sigma^2 I + \phi A^{-1} \phi^T)^{-1/2} \cdot \exp \{ t^T (\sigma^2 I + \phi A^{-1} \phi^T)^{-1} \cdot t / 2 \}
\]

(5)

Using gaussian radial basis kernel function (RBF) as kernel function:

\[
K(x, x_i) = \exp \left\{ -\frac{\|x - x_i\|^2}{2d^2} \right\}
\]

(6)

where \( d \) is the width factor of the kernel function.

### 2.2. Water quality evaluation model based on RVM

Let the time series of historical data of water quality index be \( \{y_n\}_{n=1}^N \), where \( N \) is the sequence length, \( y_n \) is the monitoring value of water quality index at time \( n \), \( x_n = [y_{n-d}, y_{n-(d-1)}, \ldots, y_{n-\tau}] \) is the vector composed of \( d \) previous monitoring values, and \( \tau \) is the sampling period. The key to establish the water quality grade evaluation model is to determine the mapping relationship:

\[
y_n = F(x), \quad n = 1, 2, \ldots, N
\]

(7)

Therefore, training sample set \( \{x_n, y_n\}_{n=1}^N \) needs to be constructed, in which \( x_n \) is the input sample and \( y_n \) is the output sample. The mapping ability of RVM model is utilized to make the input and output of RVM approximate formula (7), so as to establish the water quality classification model of RVM. On this basis, the subsequent water quality data are input into the trained RVM water quality evaluation model, and the grade evaluation of the subsequent water quality data can be completed in this model. The data obtained from the real-time system of automatic water quality monitoring were divided into three groups. Eighty percent of the data were used as training data of RVM model, totaling 3,008 groups, and the remaining 20% were used as test data of RVM model, totaling 752 groups.

Figure 1 shows the schematic diagram of water quality grade evaluation results of RVM model, and the evaluation accuracy is 81.51%. It can be seen from the figure that the evaluation errors are mainly concentrated in class III, accounting for 37.41% of the total error rate, and the main error rate is concentrated in Class II. The reason is that the actual water quality status of Class II and Class III is similar. As a result, the RVM model fails to effectively evaluate the two water quality types. Since there are altogether eight water quality evaluation factors used in this paper, there is mutual interference among different water quality evaluation factors, which leads to the situation of low evaluation accuracy.
3. WATER QUALITY EVALUATION MODEL BASED ON RELIEFF-RVM

3.1. ReliefF algorithm introduction

At present, when evaluating surface water, Chinese and foreign scholars mostly establish evaluation models directly through water quality data, often ignoring the impact of weight of each pollutant on water quality, or artificially assigning weight to each evaluation factor with insufficient consideration (Su et al. 2020). Therefore, in this paper, ReliefF algorithm was used to calculate the weight of each evaluation factor, and the evaluation factor with significant weight was selected to establish the equivalent evaluation model of water quality.

Relief is a multi-variable feature selection algorithm based on the weight of sample features proposed by Kira (Kira & Rendell 1992). As Relief algorithm is mainly applied to dichotomies, ReliefF algorithm is finally selected to complete the reduction of evaluation factors because the water quality evaluation factor involved in this paper is a multi-class selection problem. ReliefF algorithm determines the weight of samples by the difference of features between similar samples and dissimilar samples, the degree of association between features is expressed by the distinguishing ability of features to close samples to measure their classification ability (Turker et al. 2021). The specific calculation process is as follows:

Let \( X = \{x_1, x_2, \cdots, x_n\} \) be a set of samples, each containing \( m \) features, namely \( x_i = \{f_1, f_2, \cdots, f_m\} \). \( C = \{c_1, c_2, \cdots, c_l\} \) is the class set to which the sample belongs, and there are \( l \) classes in total.

1. For each feature sample, the weight \( w_i = 0 (1 \leq j \leq m) \) is initialized;
2. A sample \( x_i \) is randomly separated from \( X \), and \( d \) samples closest to \( x_i \) are selected from the samples of the same category to form the set \( H \) for the characteristic variable \( f_j (1 \leq j \leq m) \) in the sample. \( d \) samples closest to \( x_i \) are selected from each heterogeneous sample to form the set \( M(c) \), and \( 1 \leq c \leq l \) represents the class;
3. According to formula (8), the weight \( w_j \) of characteristic variable \( f_j \) is updated.

\[
\begin{align*}
    w_j & = w_j - \sum_{n=1}^{l} \frac{\text{diff}(f_j, x_i, H_n)}{t \times d} + \sum_{c=\text{class}(x_i)} \frac{p(c)}{1 - p(\text{class}(x_i))} \frac{\sum_{n=1}^{l} \text{diff}(f_j, x_i, M_n(c))}{t \times d} \\
\end{align*}
\]

where, \( p(c) \) is the probability of occurrence of class \( c \) targets.
4. Turn (2) iterates \( t (1 < t) \) times;
5. The weights of all characteristic variables are calculated, and finally the weight vector set \( W \) is obtained.
6. The elements in weight vector \( W \) are sorted from large to small, and the features larger than a certain threshold weight are selected as the target features.

3.2. Water quality evaluation network based on ReliefF-RVM

In this section, each evaluation factor is selected as the unit to calculate the characteristic weight of different evaluation factors, which is regarded as the contribution degree of evaluation factors to water quality grade evaluation. A total of eight evaluation factors of water quality were selected in this paper. Each evaluation factor was taken as an independent variable. ReliefF algorithm was used to calculate the weight of each evaluation factor.

1. The water quality parameters corresponding to each evaluation factor constitute a vector, namely: \( F = \{f_1, f_2, \cdots, f_i\} \), where \( i \) represents the parameter of the \( i \) water quality evaluation factor, \( 1 \leq i \leq 8 \);
2. ReliefF algorithm was used to calculate the feature weights of each evaluation factor, namely, \( W = \{w_1, w_2, \cdots, w_i\} \), where \( i \) represents the parameter of the \( i \) water quality evaluation factor, \( 1 \leq i \leq 8 \);
3. From the six types of water quality data, different groups of water quality data are taken, the weight of each evaluation factor is calculated repeatedly and the arithmetic average is taken, as shown in Equation (9). The average weight of each evaluation factor is obtained as the weight of this evaluation factor, and this is used as the basis for the reduction of...
evaluation factors.

\[ W = \frac{1}{N} \sum_{i=1}^{N} W_i \]  

(9)

where, \( N \) is the calculation times.

The calculation results are shown in Figure 2 through multiple calculations of data weight of evaluation factors in different basins. It can be seen from Figure 2 that different evaluation factors contribute different degrees to water quality data. The weight threshold of evaluation factors is set as, and evaluation factors lower than the threshold can be regarded as invalid factors and discarded.

According to the threshold, PH value, conductivity and turbidity can be eliminated to avoid the use of redundant evaluation factors in the subsequent grade evaluation process. Further verification is needed to further verify the reliability of the evaluation factor combination obtained by ReliefF. The purpose of reducing the evaluation factors is to achieve high grade separability with as little data as possible. As a feature of water quality grade evaluation, the evaluation factors can be used to measure the separability of different combinations of evaluation factors by means of Clustering Separation Index (CSI) (Yanzhao 2015). The steps to calculate the separability of the combination of evaluation factors through CSI are as follows:

(1) Calculate the dispersion degree \( C_{ii} \) in the class that measures the dispersion degree between the same action mode of the sample and the inter-class dispersion degree \( C_{ij} \) in the class that measures the dispersion degree between different evaluation factors. The calculation process is as follows:

\[ C_{ii} = \sqrt{\frac{1}{N} \sum_{k=1}^{N_i} \| X_k - M_i \|^2} \]  

(10)

\[ C_{ij} = \| M_i - M_j \| \]  

(11)

where, \( X_k \) is the sample in class \( i \), \( N_i \) is the number of samples in class \( i \), and \( M_i \) is the mean value of the feature vector of class \( i \).

(2) Calculate the similarity \( S_{ij} \) between the two types of samples. When \( S_{ij} \) is smaller, the separability of the two types of samples is stronger:

\[ S_{ij} = \frac{C_{ii} + C_{jj}}{C_{ij}} \]  

(12)

\[ \text{Weight of Evaluation Factor} \]

![Figure 2](http://iwaponline.com/wst/article-pdf/85/3/799/1004225/wst085030799.pdf)

Figure 2 | Schematic diagram of weight of evaluation factors of water quality evaluation.
(3) Quantify CSI index to determine the overall separability of all K evaluation factors. CSI index takes the mean value of the similarity between two types of samples that is most difficult to distinguish as the separability of the whole sample. The smaller the value is, the better the overall separability performance of the sample is:

\[ CSI = \frac{1}{K} \sum_{i=1}^{K} \max_{j \neq i} S_{ij} \]  

(13)

Figure 3 shows the CSI value of the combination of evaluation factors obtained when different evaluation factors are reduced, and the corresponding reduction grouping of evaluation factors is shown in Table 2. Because CSI value is lower, on behalf of the higher the separability of the evaluation factor combination, as shown in Figure 3, when the PH value, conductivity and turbidity reduction, its corresponding CSI value is the lowest, only 0.81, in eight kinds of evaluation factors, CSI value reached 0.99, while the continued reduction of ammonia nitrogen, CSI has increased, on behalf of its separability is reduced. Therefore, ammonia nitrogen is reserved. At the same time, the accuracy of the reduced combination of evaluation factors will be verified in the RVM model.

![Figure 3](image)

**Figure 3** | CSI value of evaluation factor combination.

**Table 2** | Combination of evaluation factors

| Category | Reduced Evaluation Factor                  |
|---------|-------------------------------------------|
| 1       | No reduction                              |
| 2       | PH value                                  |
| 3       | Conductivity                              |
| 4       | Turbidity                                 |
| 5       | PH value, conductivity                     |
| 6       | PH value, turbidity                       |
| 7       | Conductivity, turbidity                   |
| 8       | PH value, conductivity, turbidity          |
| 9       | PH value, conductivity, turbidity, NH₃-N  |
PH value, conductivity, and turbidity data in test set data and training set data were removed, and the reduced data were input into RVM water quality grade evaluation model, and the evaluation accuracy reached 83.24%. The evaluation results are shown in Figure 4. The evaluation accuracy of ReliefF-RVM after reduction is 1.73% higher than that of RVM without reduction, which verifies the effectiveness of ReliefF for water quality evaluation factor reduction scheme proposed in this section. At the same time, it can be seen that among the eight evaluation factors included in the experimental data, some evaluation factors have a negative effect on the evaluation accuracy, such as water PH value, conductivity, and turbidity. The reduction of water quality evaluation factors not only reduces the amount of experimental data of the model and improves the evaluation accuracy of the model, but also effectively reduces the workload and provides technical support for the actual water quality measurement in the process of reducing the number of measurement types of evaluation factors.

4. OPTIMIZATION OF RVM WATER QUALITY EVALUATION MODEL BASED ON PSO

4.1. Principle of PSO algorithm

Compared with neural network and support vector machine, RVM algorithm has fewer relevant vectors and fewer adjustable parameters of the model, which to some extent reduces the risk of affecting the model generalization performance due to improper parameter setting (Bishop 2021). However, the selection of parameters of the RVM model still depends on human experience. Therefore, PSO was used to optimize the parameters of the RVM model, and the kernel function parameters were optimized by PSO to build a water quality evaluation model based on PSO-RVM.

The basic concept of PSO algorithm is derived from the study of foraging behavior of birds. In the PSO algorithm, each particle represents a solution to be optimized, and the particle velocity determines the distance and direction of particle movement, and adjusts dynamically through the movement of itself and other particles, so as to realize the individual optimization process in solvable space.

PSO-RVM evaluation model construction steps are as follows:
(1) Determine the structure and related parameters of RVM model;
(2) Set the community size, initial particle velocity and corresponding point location. The optimal point of particle is the initial point, and the optimal point of community is the global optimal point;
(3) Each particle contains a fitness value, which is used to reflect the quality of the particle. After training the RVM model, the obtained training error is used as the fitness value of the particle's current point, and the result is compared with the value of the previous best point. If it is better than the previous best point, it will be replaced, otherwise it will remain unchanged;
(4) If the global optimum point is not as good as the previous optimum point of the current particle, the previous optimum point is used to replace the global optimum point, otherwise unchanged;
(5) According to Equations (14) and (15), the flight speed and corresponding point position of the re-planned particle are calculated.

\[
V_{id}^{k+1} = \omega V_{id}^k + C_1 \text{random}(0, 1)(P_{id}^k - X_{id}^k) + C_2 \text{random}(0, 1)(P_{gd}^k - X_{id}^k) \tag{14}
\]

\[
X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \tag{15}
\]

where, \( \omega \) is the inertia factor, \( C_1 \) and \( C_2 \) are acceleration constants, and are the individual learning factor and social learning factor respectively, \( \text{random}(0, 1) \) represents the random number on interval \([0, 1]\), \( P_{id}^k \) represents the individual optimal solution of the \( i \)-th particle, \( P_{gd}^k \) represents the global optimal solution, and \( X_{id}^k \) represents the position of the \( i \)-th particle;

(6) Check whether the algorithm meets the termination condition (the number of iterations reaches the maximum number of iterations or the error precision reaches the target error precision set initially). If the termination condition is true, the best weight threshold value will be output, and then the model will be further simulated; otherwise, skip to step (3).

### 4.2. Water quality evaluation network based on PSO-RVM

The construction process of PSO-RVM model is shown in Figure 5. For the particle swarm optimization network, the particle swarm size is set as 50, the inertia factor is 0.5, \( C_1 \) is 0.2, \( C_2 \) is 0.5, \( r_1 \) and \( r_2 \) are random numbers between \([0, 1]\), and the number of iterations of the particle swarm is 20. The weight and threshold of particle swarm optimization were taken as the initial parameters of RVM model to realize modeling of water quality evaluation, improve the accuracy of RVM model and reduce the randomness of system evaluation.

In order to verify the effectiveness of the PSO-RVM water quality grade evaluation model, the training set and test set used in section 2.2 were input into the PSO-RVM evaluation model to evaluate the evaluation model. The test results of the PSO-

![Figure 5 | PSO optimization of RVM network flow chart.](image-url)
RVM evaluation model were shown in Figure 6, and the evaluation accuracy reached 92.28%. The evaluation accuracy of the optimized RVM model is improved by 10.77% compared with that before optimization, which verifies the evaluation accuracy of the PSO-RVM model. At the same time, compared with the evaluation error rate of class III water quality, it was found that the evaluation error rate of class III water quality after optimization was only 3.86% of its own and 13.79% of the global evaluation error rate, which was 23.62% lower than that of class III water quality in the RVM model, which effectively verified the effective optimization of PSO for RVM network.

Meanwhile, the digital data reduced by ReliefF algorithm in section 3.2 were input into the PSO-RVM model to further verify the classification accuracy of PSO-RVM network.

Figure 7 shows the schematic diagram of water quality grade evaluation results of PSO-ReliefF-RVM model, and its evaluation accuracy reaches 95.74%. Compared with ReliefF-RVM evaluation model, it is improved by 12.5%, and compared with the error rate of class II water quality, the error rate of class III water quality evaluation after optimization is only 1.45% of its own and 9.38% of the global evaluation error rate, which verifies the optimization performance of PSO algorithm for RVM network.

Figure 8 shows the RVM model and the evaluation accuracy obtained by PSO optimization algorithm and ReliefF data reduction. As shown in Figure 8, the comparison between ReliefF-RVM model and RVM model, and between PSO-ReliefF-RVM model and PSO-RVM model shows that the water quality assessment factors reduced by ReliefF algorithm can better reflect the differences among different grades of water quality. The evaluation accuracy of RVM network is improved by 1.73% and 3.46% respectively, and the experimental data amount of RVM model is effectively reduced, which provides theoretical basis for actual production and life.
Meanwhile, comparing PSO-RVM with RVM and PSO-ReliefF-RVM with ReliefF-RVM, it is found that the evaluation accuracy of the RVM model after PSO optimization is 10.77% and 12.5% higher than that of the RVM model before optimization, which effectively verifies the optimization effect of PSO algorithm on RVM network.

5. CONCLUSION

In this paper, the RVM model was used to build a water quality evaluation model. However, due to the difficulty in determining the initial parameters of the RVM network and the wide variety of water quality evaluation factors selected in this paper, feature redundancy was caused to a certain extent. Therefore, PSO algorithm was used to optimize the RVM network. The kernel width factor and hyperparameters of the RVM model were optimized, and the experiment verified that the PSO algorithm effectively improved the accuracy of the water quality evaluation model based on RVM.

Reuse ReliefF algorithm for the reduction of evaluation factors in the eight kinds of water quality evaluation factors to cut out the PH value, ammonia nitrogen and total phosphorus. Three small weights of evaluation factors effectively reduce the amount of data used in the model and reduce the evaluation factor collection types, when using a small amount of evaluation factors, still can obtain a better evaluation of accuracy for the actual production and life to make the theoretical guidance.

ACKNOWLEDGEMENTS

This paper was supported by the science and technology development project 20190303080sf of Jilin Province.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Bishop, C. M. 2021 Pattern Recognition and Machine Learning. 2021-08-30, Springer Science, New York, NY, USA.
GB 3838-2002. 2002 Environmental Quality Standards for Surface Water. State Environmental Protection Administration, Beijing, China.
Guangyue, J. 2020 Research on water quality prediction of Xijiang River based on IMPROVED particle swarm optimization algorithm and BP neural network. Research and Progress in Hydrodynamics (Series A) 35 (05), 567–574.
Guozhong, W., Zhongyuan, L., Jiyu, Z., Qiting, Z. & Huanling, C. 2019 Analysis of influencing factors of water quality of main reservoirs in henan province based on decision tree. Journal of Wuhan University (Engineering Edition) 52 (09), 774–781.
Henghua, Z., Zhizheng, L., Chao, J., Han, Y. & Shuang, L. 2021 Application of entropy-set pair analysis method in groundwater quality evaluation in Dawu. Groundwater 43 (02), 11–16.
Hong, W., Junlan, Y., Yanqiang, L., Hsiangte, K. & Hongbing, C. 2017 Evaluation and analysis of Liangzi Lake water quality based on principal component analysis. *Journal of Hubei University (Natural Science Edition)* 39 (06), 601–608.

Jing, Y. 2014 *Application of Improved Fuzzy Comprehensive Evaluation Method in Water Quality Evaluation*. Chongqing University, Chongqing, China.

Jingsheng, C. 2000 Research progress of land water quality change. *Journal of Environmental Science* 20 (1), 10–15.

Kangyang, C. H. E. N., Hexia, C. H. E. N., Chuanlong, Z. H. O. U., Yichao, H., Xiangyang, Q., Ruoqin, S., Fengru, L., Min, Z., Xinyi, Z., Jinfeng, W., Yan, Z., Da, C., Xingguo, C., Yongfeng, D. & Hongjiang, R. 2020 Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Research* 171, 1–10.

Kira, K. & Rendell, L. A. 1992 The feature selection problem: traditional methods and a new algorithm. In: *Tenth National Conference on Artificial Intelligence*. AAAI Press, Palo Alto, CA, USA.

Lei, C. & Xiaojun, X. 2006 Application of water quality index method in raw water quality evaluation. *China Rural Water and Hydropower* 04, 8–9.

Lipin, J., Jingjing, G., Yunyun, Y., Mingmin, W., Wei, L., Xizhong, Z., Tu, S. & Meirong, M. 2020 Application of decision tree model and logistic regression model in analysis of influencing factors of drinking water quality. *Chinese Health Statistics* 37 (06), 874–877 + 882.

Ministry of Ecology and Environment. PRC. National Surface Water Quality Automatic Monitoring Real-Time Data Release System[EB/OL]. http://106.37.208.243:8068/GJZ/Business/Publish/Main.html

Nahman, J. & Salamon, D. 2017 Mutual interference of neighboring grounding systems and approximate formulation. *Electric Power Systems Research* 41 (12), 151–156.

Su, L., Zhihong, Y., Dan, X. & Shuqian, W. 2020 Application of improved Nemerow index method in reservoir water quality evaluation. *Science Technology and Engineering* 20 (31), 13079–13084.

Tengfei, M., Wenbo, C. & Yingbo, H. 2018 Application of PRINCIPAL component analysis in water quality evaluation of Gaozhou reservoir. *Sichuan Environment* 37 (01), 65–71.

Tipping, M. E. 2001 Sparse Bayesian learning and relevance vector machine. *Journal of Machine Learning Research*. 3 (1), 211–244.

Turker, T., Erhan, A. & Sengul, D. 2021 Multileveled ternary pattern and iterative ReliefF based bird sound classification. *Applied Acoustics* 17 (6), 31–39.

Weigu, C., Yabin, L., Yan, S. & Jian, W. 2019 Comparative analysis of comprehensive water quality identification index method with different weighting methods. *Journal of Irrigation and Drainage* 38 (11), 93–99.

Xiaojuan, C. 2021 *Study on environmental impact Assessment system of urban wastewater pollution treatment project*. *Energy Saving and Environmental Protection* 09, 41–43.

Xiaoyan, W., Hao, W., Jianjun, G. & Tianhong, Z. 2021 Optimization of implementation Standard of Water Quality monitoring section in Lanzhou section of Yellow River Based on cluster analysis. *Journal of Lanzhou Jiaotong University* 40 (05), 79–84.

Xueqing, L., Hang, Z., Yueyi, L. & Wenhua, W. 2021 Research on regional water Quality Prediction method based on multi-source data machine learning. *Water Resources and Hydropower Technology* 52 (11), 1–14.

Yan, Z., Zhenliang, L. & Can, Z. 2021 Study on surface water quality assessment based on decision tree support vector machine. *Environmental Science and Technology* 34 (05), 57–61 + 66.

Yang, M., Xianqiao, J., Yuhui, Z., Lihua, Z. & Chengbo, F. 2021 Evaluation of groundwater quality based on improved matter-element extension model. *Journal of Hebei Geological University* 44 (04), 90–98.

Yanzhao, C. 2015 *Motion Pattern Recognition of Autonomous Rehabilitation Training of Upper Limbs Based on sEMG Signal*. Shandong University, Ji Nan, Shandong, China.

Yongcai, R., Yuan, Y., Ting, S., Wei, P., Xiaowei, Z., Yafeng, L., Yujing, X. & Zhengqiu, F. 2021 Water quality evaluation and pollution source analysis of Beijing-Hangzhou Canal (Xuzhou section). *Journal of Fudan University (Natural Science)* 60 (05), 596–606.

Yongjun, Z., Qiong, W. & Zhongyu, Z. 2021 Water quality evaluation based on principal component analysis and artificial neural network coupling model. *Jiangsu Water Resources* 08, 48–54.

Yuanyuan, M., Jingjing, Y., Jinsong, Z., Huan, W. & Chaojun, G. 2015 Application of Delphi method in constructing comprehensive evaluation index system of water quality automatic monitoring station. *Green Technology* 07, 180–182.

Zhijun, H. 2020 River water quality evaluation model based on fuzzy comprehensive matrix. *Hydraulic Science and Cold Region Engineering* 3 (04), 158–161.

First received 30 September 2021; accepted in revised form 14 December 2021. Available online 10 January 2022.