Water Science & Technology

IWA PUBLISHING

Water Science & Technology Vol 88 No 8, 2108 doi: 10.2166/wst.2023.310

Model evaluation of total phosphorus prediction based on model accuracy and interpretability for the surface water in the river network of the Jiangnan Plain, China

Hao Zhang[a], Juan Huan[a,*], Xiangen Xu[b], Bing Shi[a], Yongchun Zheng[a], Jiawei Mao[a] and Jiapeng Lv[a]

[a] School of Computer and Artificial Intelligence, School of Alibaba Cloud Big Data, School of Software, Changzhou University, Changzhou 213100, China
[b] Changzhou Environmental Science Research Institute, Changzhou 213002, China
*Corresponding author. E-mail: huanjuan@cczu.edu.cn

ABSTRACT

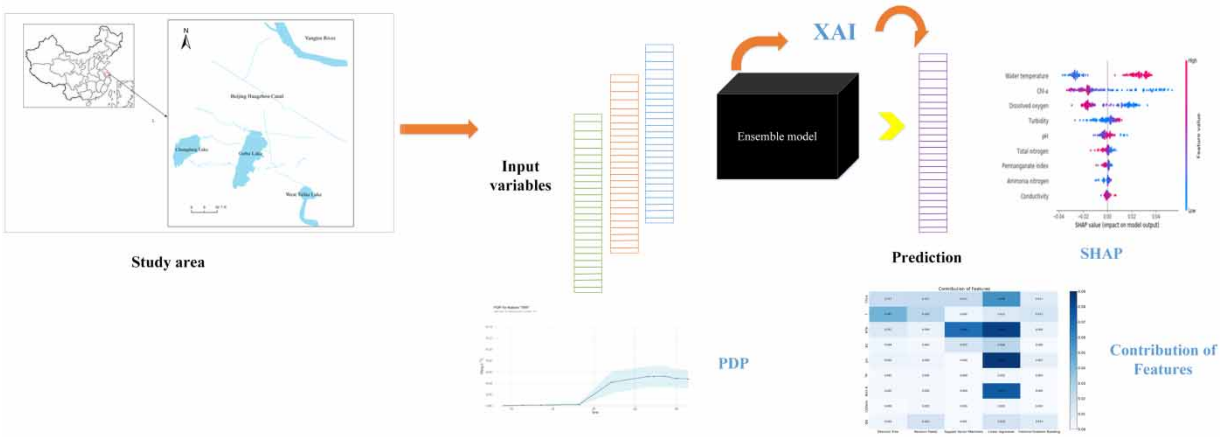
Due to climatic and hydrological changes and human activities, eutrophication and frequent outbreaks of cyanobacteria are prominent in the Jiangnan Plain basin of China. Therefore, building a suitable model to accurately predict the phosphorus concentration in surface water is of practical significance to prevent the above problems. This study built 10 models to predict the phosphorus element in the surface water of the river network in the Jiangnan Plain. The main water types in the basin include the Yangtze River, the Beijing-Hangzhou Canal, and the Gehu Lake. The 10 models in different datasets have been comprehensively evaluated by the prediction accuracy and interpretability of the model, and the calculation of the partial dependence diagram (PDP) and SHAP has proved that there is a transparent response relationship between phosphorus and different factors. The results show that the Yangtze River, Beijing-Hangzhou Canal, and Gehu Lake are suitable for random forest, linear regression, and random forest models, respectively, under the comprehensive evaluation of the prediction accuracy and interpretability of the model. Models with low prediction accuracy often show strong interpretability. In different water body types, turbidity, water temperature, and chlorophyll-a are the three factors that affect the model in predicting phosphorus.

Key words: model evaluation, model interpretability, PDP, SHAP, total phosphorus prediction

HIGHLIGHTS

- Construct 10 models based on three datasets: the Yangtze River, the Beijing-Hangzhou Canal, and the Gehu Lake.
- Three criteria for model interpretability were proposed and 10 models were ranked for interpretability.
- It was found that water temperature, chlorophyll-a, and turbidity were the most influential factors in predicting total phosphorus in the three water quality datasets.

GRAPHICAL ABSTRACT



INTRODUCTION

According to the 2021 Bulletin on the State of China's Ecology and Environment, 15.2% of the 3,632 surface water sections monitored in China are below class III water standards. The main pollution indicators are total phosphorus, chemical oxygen demand, and permanganate index (China 2022). Phosphorus is the most important nutrient element affecting the nutrient status and phytoplankton productivity in natural lakes and reservoirs and is the most concerned element in the ecological environment protection of lakes and rivers (Elser & Patrick 1994). The middle and lower reaches of the Yangtze River is China's most important urbanized regions and economic core area. With the development of agriculture and urbanization, the lake water body eutrophication and algal blooms is increasingly serious. The fundamental reason is the increase of nutrients, mainly for phosphorus and nitrogen (Reichwaldt & Ghadouani 2012; Xia *et al.* 2020). However, there are great challenges in predicting total phosphorus in watershed lakes because the close relationship between phosphorus elements and neglected hydrological, hydrodynamic, and meteorological conditions in these regional waters was ignored. Previous studies on phosphorus mainly focus on exploring the physical, chemical, and biological behaviors of phosphorus in surface water (Xu *et al.* 2015; Zhu *et al.* 2015), but there are relatively few studies on high-frequency monitoring data in watersheds. Therefore, there is still a need for frequency and high-speed online monitoring systems, as well as integrated advanced modeling methods (Thomas *et al.* 2018; Thomson-Laing *et al.* 2020). In the construction of models, it is necessary to not only pay attention to the prediction accuracy but also to obtain the intrinsic trends and potential relationships of phosphorus from the complex watershed level, which requires the integration of advanced modeling and analysis methods (Koch *et al.* 2001; Barzegar *et al.* 2020).

In recent years, more and more scholars have modeled and predicted phosphorus in watershed surface water and evaluated its predictive performance (Chen *et al.* 2015; Recknagel *et al.* 2017; Tong *et al.* 2019). Water resource managers and government departments are also slowly accepting predictive modeling tools in the field of machine learning. However, during the use of these modeling tools, the models were found to have an overall black box nature, resulting in unclear internal relationships in the prediction of phosphorus in surface water. Through further discussion, the question was raised at the current stage: can the model show good predictive performance while maintain high quality interpretability? At the same time, we are also thinking about the form in which the model can show its internal interpretability, such as neural network, which has a good prediction effect on time series data but cannot intuitively show the response relationship between the total phosphorus of surface water and its characteristics (Breiman 2001; Donoho 2017). In this study, we used the interpretability of the model and the prediction accuracy as the model of two evaluation standards. To this end, we chose three different types of water datasets in the middle and lower reaches of the Yangtze River delta plain river network region and contrasted 10 different machine learning models.

Ranking the 10 models on the dataset in terms of prediction accuracy and interpretability, it is clear that the ranking of prediction accuracy is quantitative, while the ranking of interpretability is qualitative. Based on the data from these two rankings, taking into account the model's prediction accuracy and interpretability, we would like to know if we can obtain predictive models suitable for phosphorus elements in the Yangtze River, the Beijing-Hangzhou Canal, and the Gehu Lake. Therefore, the objectives of this study are as follows: (1) to establish the interpretable ranking based on the water phosphorus prediction model in Jiangnan Plain; (2) determine the total phosphorus prediction model suitable for three water bodies, including the Yangtze River, the Beijing-Hangzhou Canal and the Gehu Lake; and (3) the transparent response relationship between surface water phosphorus element and various factors was obtained.

RESEARCH AREA DATA AND METHODS

Research area and data source

The river network area of the Jiangnan Plain refers to the densely covered area of the middle and lower reaches of the Yangtze River in China, which has flat terrain, vertical, and horizontal rivers and highly developed river network (Liu 2020). Affected by tides, the water situation is complicated, and the boundary of the river system is blurred. Gehu Lake is an important catchment area in the upper reaches of Taihu Lake basin with typical river network characteristics of the Jiangnan Plain, as shown in Figure 1. Its average inflow accounts for 20% of the total inflow of Taihu Lake (Cheng 2022). The annual average temperature, precipitation, and wind speed (WD) of the Tao watershed from 1984 to 2021 are 16.49 °C, 1,222.44 mm and 2.577 m/s, respectively. The Gehu watershed is responsible for communicating with the Yangtze River

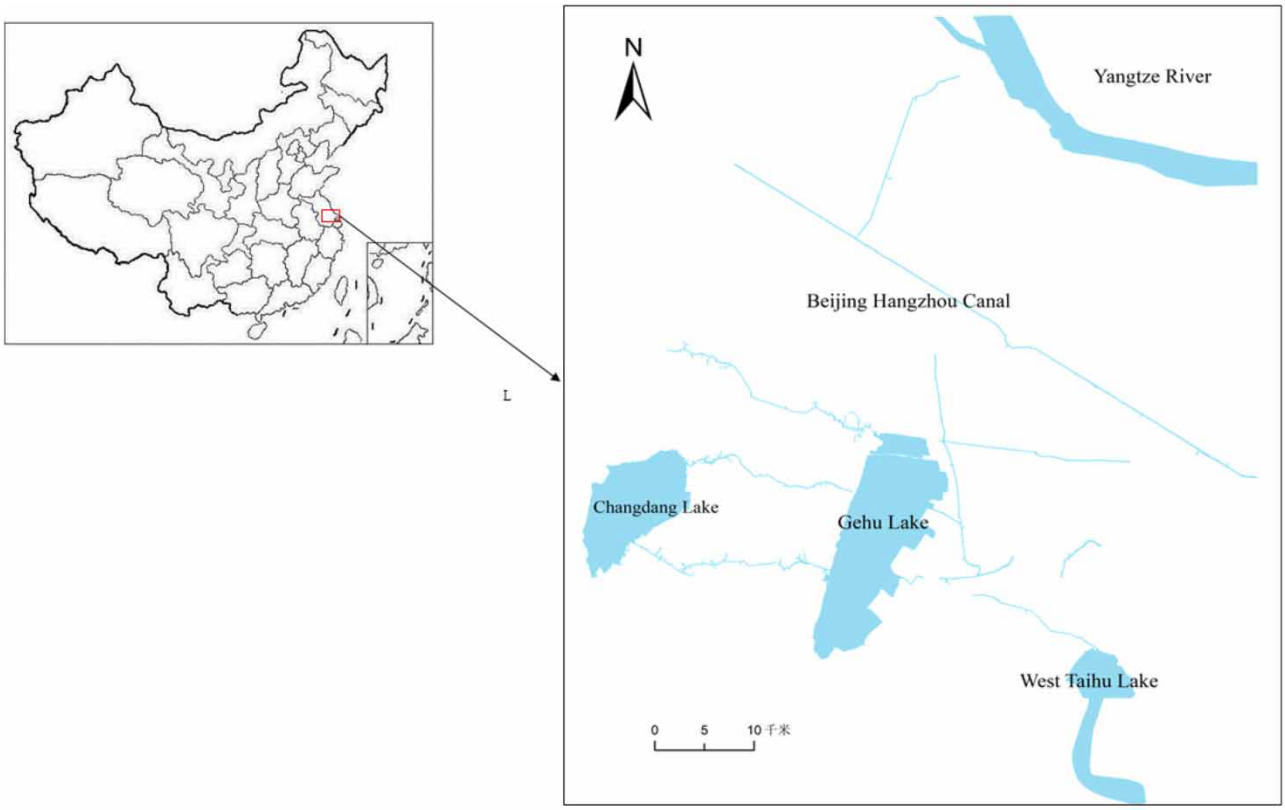


Figure 1 | River system diagram of the middle and lower reaches of the Yangtze River.

and Taihu Lake, and has the important functions of water regulation, flood control and water transportation. Since 2021, with the opening of the Xin Meng River diversion project, the Gehu watershed has undergone great changes in the hydrological and water resources pattern, such as the decrease of stagnant flow and reverse flow in the current direction and the rapid increase of the flow velocity (FV). The changes of hydrological pattern will also affect the water quality changes of the river tributaries along the process (Huang *et al.* 2020). In addition, due to the increase of urbanization and industrialization in this region, the river system also shows different degrees of eutrophication (Wu *et al.* 2022). Phosphorus is the main nutrient element affecting the nutrient status and phytoplankton productivity in natural lakes and reservoirs. It is also one of the main pollution factors in the Tao and Gehu watershed (Chao *et al.* 2023). The excessive concentration of phosphorus will not only accelerate the eutrophication of the watershed, but also destroy the ecological balance of the water (Sondergaard *et al.* 2001; Qin *et al.* 2020). Therefore, predicting the degree of eutrophication and phosphorus concentration in water bodies has become the demand of the government management department.

The work is based on several different datasets that comprise three water groups: (1) the Yangtze River, (2) Gehu Lake (large shallow lake), and (3) Beijing-Hangzhou Canal. Multiple water quality station sections are distributed in the basin, including multi-parameter sensor data acquisition, data transmission by buoy system, and data storage and processing by data center platform. Water quality monitoring data include water temperature, dissolved oxygen, pH value, conductivity, turbidity, permanganate index, total phosphorus, etc. Several meteorological monitoring stations were distributed around the basin, recording WD, air temperature (AT), and rainfall data. In these datasets, hydrologic features include FV and water level (WL). The data interval in this study is 4H, and such high-frequency monitoring data are rarely perfect in reality. Due to the fault and deviation of multi-parameter sensors and the abnormal situation of watershed water, there are inevitably missing and outliers. Therefore, based on the Chinese standard of surface water quality, this study set different effective value ranges for different indicators (except pH value, the upper bound of effective value is 50% of the worst class V water quality standard) to exclude outliers, as shown in Table 1. Therefore, the site monitoring data not within the above valid value range will be considered outliers, and will be regarded as missing values in the subsequent processing process. Before constructing

Table 1 | Basin water quality data and effective value range

Water quality index	Valid value (lake)	Valid value (river)
Water temperature (°C)	Weekly change range $-3 \sim 1.5$	
Dissolved oxygen (mg/L)	0–12	0–12
pH value	0–14	0–14
Permanganate index (mg/L)	0–22	0–22
Total phosphorus (mg/L)	0–3	0–3
Total nitrogen (mg/L)	0–0.3	0–0.6
Ammonia nitrogen (mg/L)	0–3	0–3

the prediction model, the data completion based on the random forest algorithm in Python is used to deal with this problem (Kuhn & Johnson 2016).

Multi-feature model and predictive power

Reliable water quality model is of great significance for predicting the trend of water quality change, taking control measures and establishing decision and early warning mechanism. Most of the prediction models selected in this study are from the field of 'machine learning', and some are from the neural network algorithm in 'deep learning', which has been widely used in many research fields, including water quality modeling (Yarkoni & Westfall 2017). This wide application may be due to its good properties in nonlinear simulation and excellent prediction performance, and also related to the recent success of deep learning neural network algorithms in natural language processing and pattern recognition.

As mentioned in the introduction, the 10 modeling tools selected in this study can obtain the final prediction results by learning the mapping function. Function $f(\theta)$ acts on the meteorological and water quality factor data in the T_h period, namely: $\mathrm{SWEQS}_i = [y^i_{\tau+1}, y^i_{\tau+2}, \cdots, y^i_{\tau+T_p}] = f([X_{\tau-T_h+1}, X_{\tau-T_h+2}, \cdots, X_\tau]; \theta)$. In the formula, SWEQS is short for surface water quality and environmental quality standard. The predicted water quality index selected here is total phosphorus. $y^i_{\tau+1}, y^i_{\tau+T_p}$ is the predicted value of total phosphorus in surface water of watershed section I in interval $\tau+1, , \tau+T_p$; X_τ is the physicochemical index of watershed water affecting the predicted water quality total phosphorus at time τ, including hydrological, water quality and meteorological factors, index i is the number of water body datasets in the basin. The selected 10 predictive modeling tools were all used as multi-feature models, that is, different characteristic factors were used as model inputs to output the predicted value of phosphorus element in surface water.

The prediction model selected in this study is shown in Table 2, including model types, advantages, and disadvantages. The model in the table is implemented several times to explore the optimal implementation method and hyperparameter selection. Models are listed in Table 2 by the python language environment gets its implementation.

Criteria for model interpretability

Machine learning shows great potential in the prediction process of certain fields (Doshi-Velez & Kim 2017), but computers usually do not interpret their prediction results, which need to be supplemented by the prior knowledge of experts. At present, there are several ways to explain the role of machine learning prediction in this technical process. Molnar (2021), in his book *Interpretable Machine Learning*, shows that model structures and prediction formulas for linear regression, multiple regression, and decision trees are transparent. In addition to linear regression, multiple regression and decision tree, other models have more black box characteristics, although different models show significant differences in the characteristics of the prediction process. Multi-layer perceptron (MPL) belong to the category of neural network model, which greatly facilitates us to re-write the propagation strategy based on neural network into the nonlinear response equation. Random Forest and Extreme Gradient Boosting are formed based on multiple decision trees, each of which has its transparent structure. However, if the whole 'Forest' is predicted, the interpretability will become very difficult because it cannot explain the role of different decision trees in the prediction process. Support Vector Machines are the only linear models that can classify data that cannot be linearly separated. They can interpret high-dimensional feature spaces with their specific kernels. Place the remaining three selected models on the model with low transparency, they are, respectively, Integrated Approach, Multilayer Perceptron, and long short-term memory (LSTM). Due to the fact that Multilayer Perceptron and LSTM belong to the

Table 2 | Model types and their pros and cons

Model type	Advantages	Shortcomings
Decision Tree	Easy to understand and explain	Instability and overfitting
Random Forest	Feature importance ranking	Not friendly to low-dimensional data
Support Vector Machines	Unique kernel function	Too large number of features requires regularization
Linear Regression	Fast and interpretable	Does not fit nonlinear data well
Extreme Gradient Boosting	High precision and flexibility	The sorting process has high space complexity
Nearest Neighbor	Simple and effective suitable class-domain cross-sample	Large amount of calculation and poor interpretability
Integrated Approach	Mixed data processing capability and strong robustness	Poor scalability
Multiple Regression	Simple and effective	Not very accurate and does not fit nonlinear data well
Multilayer Perceptron	Fast training and real-time learning	Not suitable for large-scale data
Long Short-Term Memory Neural Network	Gradient vanishing and memory decay in time series data	High time complexity

category of neural network models, and as more and more deep network structures with weighted parameters emerge, it is necessary to understand and interpret the predictions of such neural networks by understanding the weights of millions of parameters, which is very difficult. Therefore, we place them at the low end of model transparency (Hans *et al.* 2022).

Partial dependence plot

The transparency of different models is different, but how to understand the internal prediction process of the model is crucial. Specifically, in the process of predicting phosphorus in water bodies within the model, the response relationships between various water quality indicators and phosphorus elements are obtained. The more in line with the actual situation and expert prior knowledge, the easier the model is to be understood and the stronger its interpretability. The importance of this relationship is self-evident, perhaps second only to the predictive accuracy of the model itself. Clearly, the greatest advantage of linear regression models is linearity, which simplifies the prediction process. Most importantly, it makes the learned linear relationships easier to interpret through the weight of input features (Kuhn & Johnson 2016). Molnar (2021), discusses a general approach that attempts to simulate any machine learning model for response relationships between different features in the prediction process by computing partial dependence graphs (partial dependence plot (PDP)). The PDP can show the average marginal effect of one or two features on the prediction results of the machine learning model. It also can show whether the relationship between the surface water phosphorus element and the feature X_τ is linear, monotonic, or more complex. Therefore, PDP can reveal the correlation between surface water phosphorus element and characteristic factors $X_1, X_2 \cdots, X_M$ in the model prediction process, which may violate the relationship believed by experts' prior knowledge. The calculation of the PDP is intuitive, and if the feature of the calculated PDP is not related to other features, the PDP can perfectly represent the influence of the feature on the target prediction. Unfortunately, PDP also has a disadvantage that it can show misleading relationships if the other features in the model are strongly correlated with each other, which needs to be noted.

Shapley Additive exPlanations

Shapley Additive exPlanations (SHAP) is a method to explain individual prediction. It calculates the Shapley value of each feature based on game theory, and then explains the contribution of each feature to the final prediction value. It can make us intuitively know which features have positive or negative effects. The above two methods can provide an explanation scheme for almost all machine learning and deep learning, including tree model, linear model and neural network model. Most importantly, the interpretation of Shapley values can be visually represented as features similar to linear models.

Based on the above proposed method, we propose to apply three criteria to measure the interpretability of the model, namely: (1) the transparency of the model structure; (2) the accurate response relationship in the process of model prediction; and (3) accurate feature importance ranking (compared with the prior knowledge of experts).

RESULTS AND DISCUSSIONS

Model prediction accuracy ranking

In this study, 10 models have been developed to predict phosphorus concentration in the surface water of the Yangtze River, the Beijing and Hangzhou canals and Gehu Lake. The predicted process has repeated 10 times for each model, and the average of the results is used as the final result. The accuracy of the model is shown in Table 3, RMSE is the average root of root mean square error. The smaller RMSE is, the higher the prediction accuracy is, RMSE formula is shown in the following: $\mathrm{RMSE} = \sqrt{1/Te \sum_{k=1}^{Te} (y_k^i - Y_k^i)^2}$, in the formula, Te is the sample size of the test set; y_k^i and Y_k^i are the predicted values of lake and river i and the actual observed values of the river, respectively.

R^2 is a statistical indicator of the close correlation between response variables. More R^2 value close to 1 corresponds to higher model prediction accuracy. R^2 formula is shown in the following: $R^2 = 1 - \left(\sum_{k=1}^{Te} (y_k^i - Y_k^i)^2\right) / \left(\sum_{k=1}^{Te} (Y_k^i - \bar{Y}_k^i)^2\right)$, in the formula, $\bar{Y}_k^i$ is the average of the actual observed values of the predicted lake and river i. In the Yangtze River dataset, the prediction accuracy of LSTM Neural Network was the highest ($R^2 = 0.905$), followed by Random Forest ($R^2 = 0.894$). Due to the fact that LSTM belongs to the category of deep learning models, and there are many available data in the Yangtze River dataset, a set of data is generated every 15 min at the Yangtze River water quality monitoring station, which contains a total of 35,040 sets of data. As is well known, deep learning models can have good prediction performance based on a large dataset. The LSTM model has inherent advantages in predicting time series data, as it can effectively capture the connections between long sequences and alleviate the phenomenon of gradient vanishing or exploding. Therefore, the model achieved the best prediction accuracy in the Yangtze River dataset (Ouyang *et al.* 2021). In the Gehu dataset, the R^2 of Random Forest reached 0.862, higher than that of the LSTM Neural Network (0.853), which is the highest accuracy of the 10 prediction models. In the dataset of Beijing-Hangzhou Canal, the prediction accuracy of Random Forest, LSTM Neural Network, and linear regression ranked the top three, respectively. The reason is that random forest models can also effectively run on larger datasets, handle higher latitude data, and identify the most important features from the training dataset. This is one of the most important reasons for the high prediction accuracy of random forest models (Yang *et al.* 2022). In terms

Table 3 | Prediction accuracy of each model in the three datasets

	R^2			RMSE		
Model	Yangtze River	Beijing-hangzhou Canal	Gehu Lake	Yangtze River	Beijing-hangzhou Canal	Gehu Lake
Decision Tree	0.831	0.842	0.798	0.0017	0.0016	0.0021
Random Forest	0.894	0.878	0.862	0.0016	0.0018	0.0018
Support Vector Machines	0.795	0.815	0.782	0.0019	0.0018	0.0020
Linear Regression	0.767	0.864	0.797	0.0024	0.0017	0.0019
Extreme Gradient Boosting	0.802	0.814	0.803	0.0018	0.0018	0.0019
Nearest Neighbor	0.837	0.865	0.814	0.0017	0.0016	0.0019
Integrated Approach	0.846	0.851	0.803	0.0017	0.0017	0.0019
Multiple Regression	0.793	0.796	0.793	0.0021	0.0020	0.0019
Multilayer Perceptron	0.876	0.853	0.824	0.0015	0.0016	0.0017
Long Short-Term Memory Neural Network	0.905	0.876	0.853	0.0014	0.0018	0.0018

of prediction accuracy, the LSTM model is the best model for predicting phosphorus elements in the Yangtze River; Random Forest has the highest prediction accuracy for phosphorus elements in the Beijing-Hangzhou Canal and Gehu Lake.

Verification of model interpretability

In Section 2.3, model interpretability is judged according to three criteria. First, the models were ranked and explained based on the structural transparency of the models. Second, the relationship between the characteristic factors selected by each model and the response of phosphorus element was obtained by calculating the partial dependence diagram (PDP). In the discussion section, we introduce expert prior knowledge and physical knowledge of the actual situation of lakes and rivers to verify whether the results presented by the model match the actual situation. Finally, SHAP is used to obtain the importance ranking among features of each model in the prediction process.

The transparency of Linear Regression, Multiple Regression and Decision Tree models is the highest among the selected models. Linear Regression and Multiple Regression models obtain the target prediction value by weighting the weight relationship of each feature. The most important thing is that Linear Regression obtains the weight of the feature, which is easy for people to understand and explain intuitively. Decision Tree is a Tree with Decision rules, which can visualize the Tree structure and Decision criteria in the prediction process. The transparency of Random Forest and Extreme Gradient Boosting is second only to linear regression and decision tree. Because they are based on the structure of decision tree, they are complex representations of decision tree. The interpretability of support vector machines is a close second. Multilayer Perceptron and LSTM Neural Network belong to deep learning models. With the passage of time, more and more deep network structures with weight parameters appear. Therefore, to understand and explain the prediction of Neural Network, the weight of millions of parameters interacting in complex ways must be taken into account, which requires bringing in specific data to judge. We put Integrated Approach and Nearest Neighbor in the lowest part of the transparency of the above model. In general, linear regression and multiple regression have transparent functions. The Decision Tree can know the response relationship of different features in the prediction process and the decision based on which features are made in the model. The transparency of the model is one of the criteria to judge the interpretability of the model. In addition, it is also necessary to judge whether the feature importance and response relationship conform to the cognition of the actual situation.

In a partial dependency diagram, the relationship between one or two features and the target feature is usually shown simultaneously. Taking Gehu Lake, a typical shallow lake in the middle and lower reaches of the Yangtze River, we have constructed the Gradient Boosting algorithm, Decision Tree, Random Forest, and Linear Regression model to predict the concentration of phosphorus element in the surface water of Gehu Lake. Furthermore, PDP is used to visualize the relationship learned in the model prediction process, and the influence of different features on phosphorus in the model prediction process is obtained. Water temperature and total nitrogen are selected here, as shown in Figures 2 and 3.

During the process of forecasting with the Gradient Boosting model as shown in Figure 2(a), the combined temperature of Gehu Lake has increased by 0.058 mg/L in the range of 18–21.8 °C, and the combined process of environmental oxidation shows that the concentration of phosphorus has a slight decrease of 0.0085 mg/L after 28.2 °C. In the other water temperature range, phosphorus concentration almost did not change. In Figure 2(b), in the prediction process of Decision Tree, in the range of 14–18.1 °C, phosphorus rose slightly, margin of 0.025 mg/L, similar to Gradient Boosting, phosphorus increased by 0.065 mg/L in the range of 18.2–22.1 °C. In the Random Forest model prediction process in Figure 2(c), there is not only an increase of 0.042 mg/L in the range of 18.2–22.1 °C, but also a tiny increase of 0.012 mg/L in the range of 22.1–27 °C, which is different from the above two models. In the Linear Regression prediction process shown in Figure 2(d), phosphorus concentration increases with the increase in water temperature. When the water temperature is 30 °C, phosphorus concentration increases by about 0.016 mg/L, which is not a significant increase compared to the previous models. From these four models, we can get that the temperature of water between 18 and 23 °C has the most significant influence on the phosphorus prediction results in the process of model prediction. In shallow lakes, there is a significant correlation between meteorological factors and total phosphorus, and the total phosphorus concentration increases obviously with the temperature increase. As the temperature of lake water rises, cyanobacteria bloom erupts. More importantly, the WD disturbs the water, causing phosphorus from the sediment to be released into the water (Li *et al.* 2022; Liu & Zhang 2022). Eventually, it increases the concentration of phosphorus in the water. The relationship between the phosphorus concentration and the change in water temperature has been analyzed in the real dataset of Gehu Lake, with a positive correlation coefficient of 0.757. Further analysis shows that when the total phosphorus increases by 0.01 mg/L, the average water temperature increases by 5–8 °C,

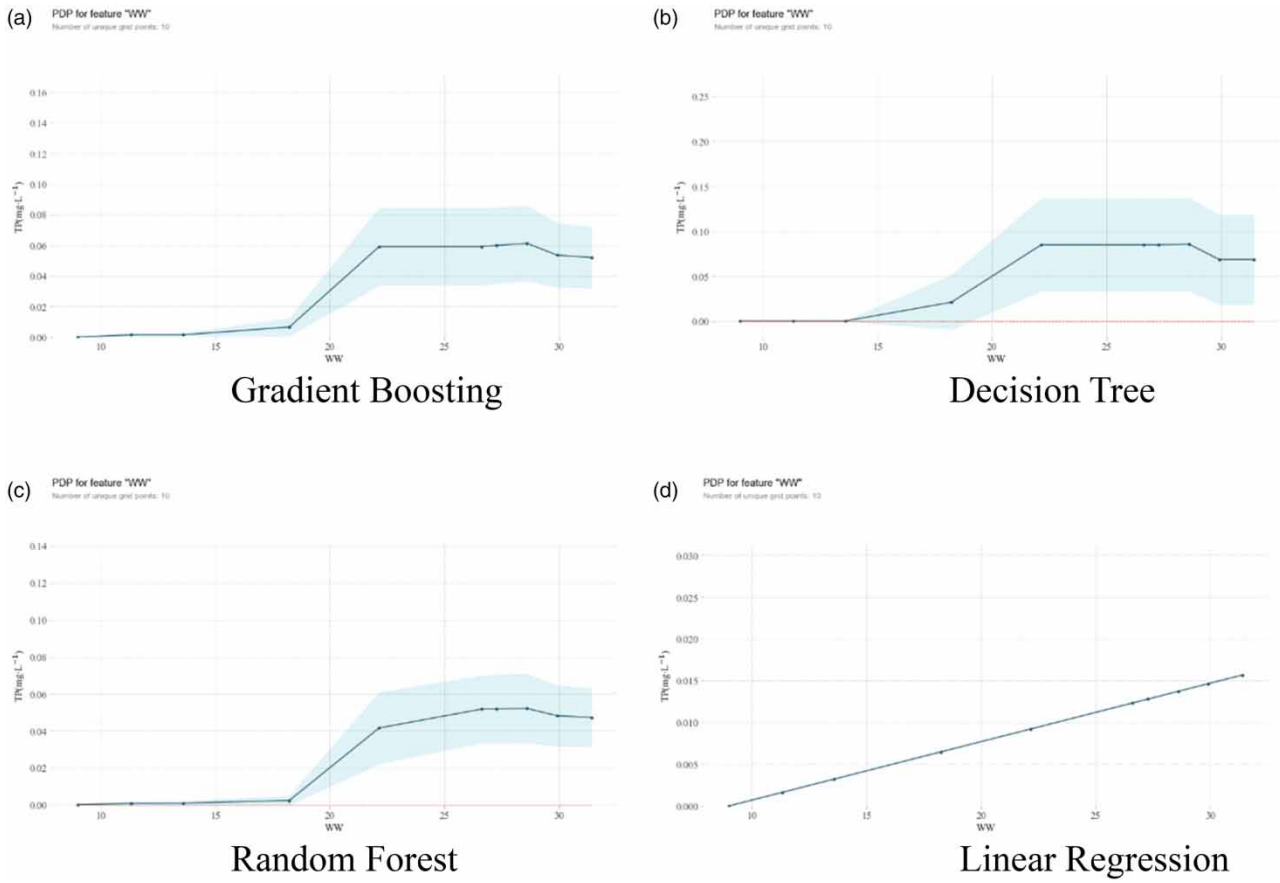


Figure 2 | PDP relationship between water temperature and phosphorus element in predicting gradient boosting, decision tree, random forest, and linear regression.

and when the temperature exceeds 15 °C, the change rate will accelerate It is most consistent with the results of the linear regression model prediction process, followed by Random Forest and Gradient Boosting, and finally, the Decision Tree.

Figure 3 shows the response relationship between nitrogen and phosphorus elements in the prediction process of each model. According to Figure 3(a), in Gradient Boosting, the combined concentration of nitrogen has increased by 0.008 mg/L in the range of 1.4–1.69 mg/L, and the combined concentration of 1.69–2.14 mg/L has slowly decreased with a decrease of 0.012 mg/L. It is worth noting that 2.11 mg/L is a critical point for nitrogen. When the concentration is less than 2.11 mg/L, nitrogen plays a positive role in predicting phosphorus. While when the concentration is greater than 2.11 mg/L, it plays a negative role in predicting phosphorus. In the process of Decision Tree prediction in Figure 3(b), the change range of nitrogen element to phosphorus element is small. In the range of 1.40–2.09 mg/L, phosphorus element increases by about 0.0015 mg/L. Similarly, in the range of 2.09–2.15 mg/L, phosphorus element concentration decreases slightly by 0.007 mg/L. Nitrogen at 2.10 mg/L is a critical point. In the Gradient Boosting and Decision Tree models, the response relationship between nitrogen and phosphorus elements is similar. In Figure 2(c) Random Forest, phosphorus increased by about 0.003 mg/L in the range of 1.40–1.68 mg /L nitrogen, and decreased in the range of 1.68–2.4 mg /L phosphorus, with different decreasing ranges in each plot, and the overall decreasing range was about 0.006 mg/L. In the Linear Regression prediction process in Figure 2(d), nitrogen negatively correlated with phosphorus. In the range of nitrogen from 1.4 to 2.4 mg/L, phosphorus decreased by 0.008 mg/L, which was higher than the previous three models. Nitrogen and phosphorus are the main influencing factors of lake eutrophication. The Gehu Lake process shows a negative correlation between nitrogen and phosphorus with a correlation coefficient of −0.173. Based on the combination of the real dataset and the prior knowledge of experts, it is estimated that the combined oxidation of 1.8 mg/L nitrogen element, when the phosphorus element decreases by 0.002 mg/L, the nitrogen element increases by 0.16–0.24 mg/L (Wu *et al.* 2018; Zhang *et al.* 2022).

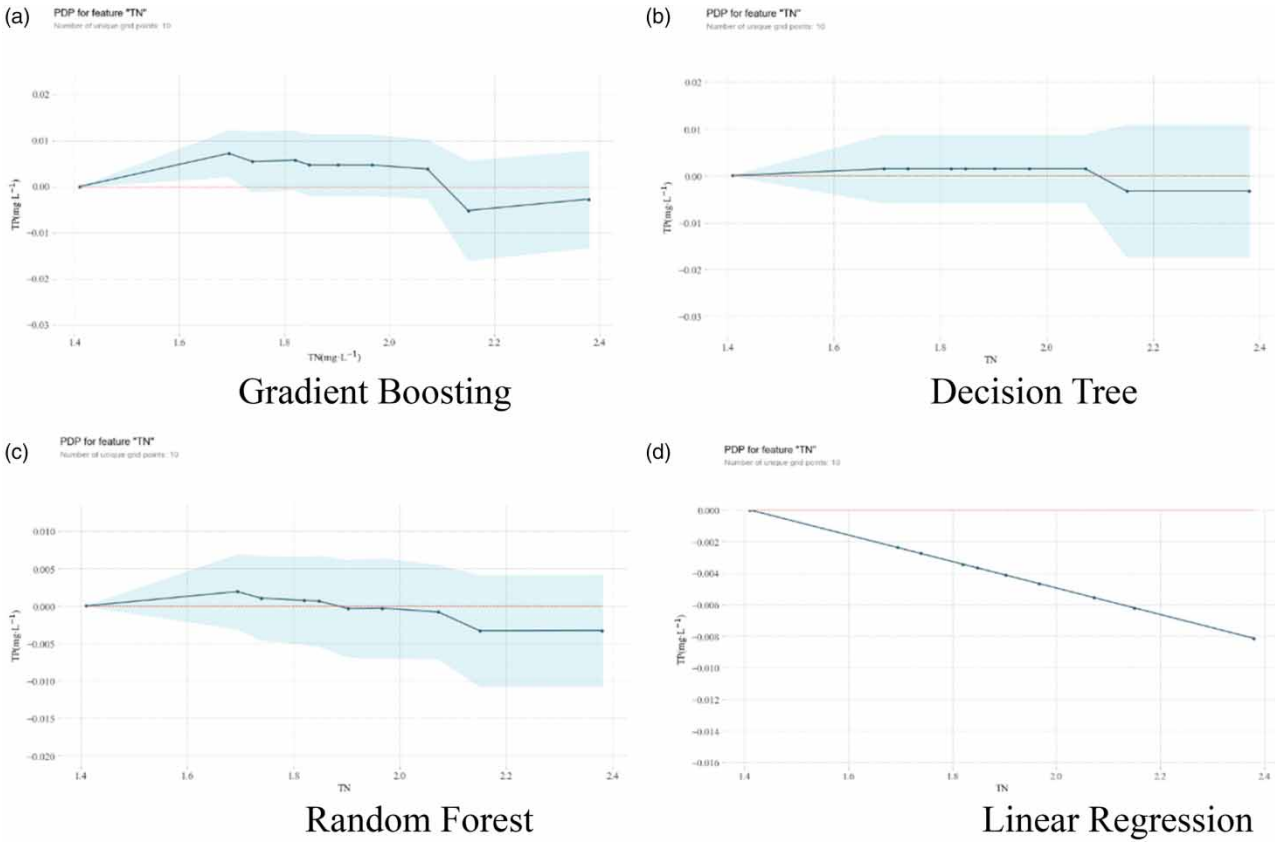


Figure 3 | PDP relationship between total nitrogen and phosphorus in the prediction process of gradient boosting, decision tree, random forest, and linear regression.

Based on the third criterion of interpretability, that is, feature importance ranking. In the process of model prediction, Tree SHAP and Kernel SHAP were used to obtain the change characteristics of phosphorus elements with different characteristic variables and rank them. Taking the datasets of Gehu Lake and the Beijing-Hangzhou Canal as an example, we used the Tree SHAP method to construct the Gradient Boosting algorithm, Decision Tree, and Random Forest model. The Support Vector Machines and Linear regression models were constructed by the Kernel SHAP method. Based on this, the influence of different characteristics on phosphorus in the process of model prediction can be quantitatively obtained. It is worth noting that the influence size here is not equivalent to the causal relationship between the actual water characteristics, but only reflected in the model prediction process.

Figure 4(a)–4(e) shows the influence of Gradient Boosting, Decision Tree, Random Forest, Support Vector, and Linear regression on phosphorus in the prediction process, respectively. It can be seen that in the prediction process of Gradient Boosting, Decision Tree, and Random Forest models, the higher the water temperature, the higher the predicted value of phosphorus, which has a positive effect on phosphorus. It can be seen that, on the contrary, the lower the concentration of chlorophyll-a, the lower the predicted value of phosphorus. The Support Vector Machines showed that turbidity changed most obviously for phosphorus, and the higher the turbidity, the higher the predicted value of phosphorus. According to Figure 5, Gradient Boosting, Decision Tree and Random Forest are all constructed based on Tree SHAP, so they are compared together. In the Gradient Boosting prediction process, the phosphorus element change contributions of water temperature, chl-a, dissolved oxygen and pH are 0.016, 0.011, 0.010, and 0.007, respectively, and the units are mg/L. In Decision Tree, water temperature, chl-a, turbidity, and pH ranked the top 4, with the contributions of 0.043, 0.021, 0.012, and 0.005, respectively. Random Forest is similar to Gradient Boosting, which is water temperature, chl-a, dissolved oxygen and pH, and the variation values of phosphorus element are 0.026, 0.022, 0.016 and 0.0058, respectively. Support Vector Machines and Linear Regression were built based on Kernel SHAP. In the prediction process, turbidity on phosphorus element

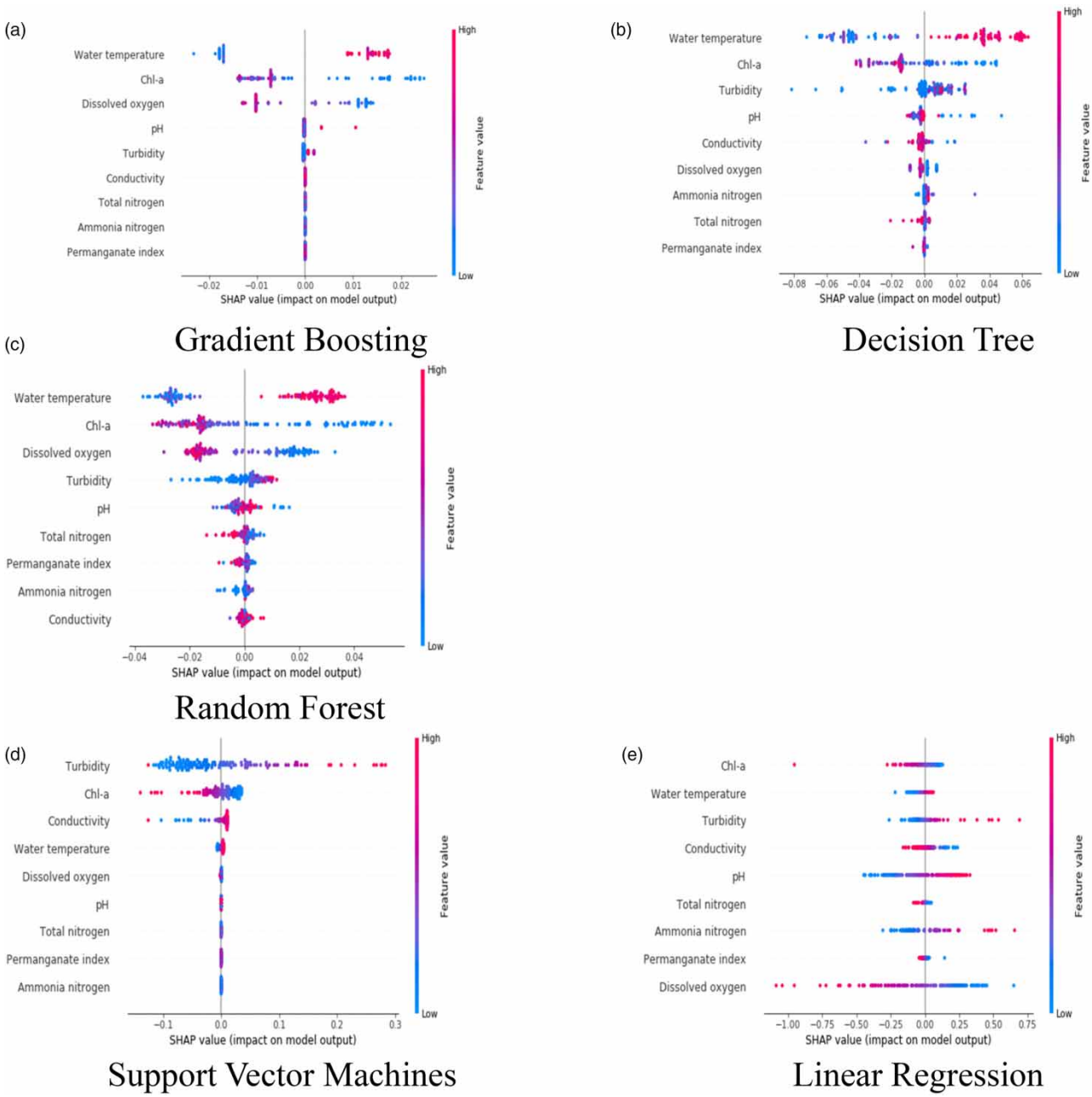


Figure 4 | Response relationship between water quality factors to phosphorus in the prediction process of gradient boosting, decision tree, random forest, support vector machines, and linear regression.

changed the most, reaching 0.068. chl-a and conductivity close behind at 0.022 and 0.015, respectively, and water temperature was 0.004. It is important to note that the characteristics here are not always positive correlation, but also negative correlation.

The explanatory ranking of the models based on the three explanatory criteria above is shown in Table 4. First, the linear regression, multiple regression, and decision tree models have the best interpretability and are classified in the first category. Next, the Random Forest, extreme gradient random trees, ensemble learning algorithms, and support vector machines have the best interpretability and are therefore classified as the second category. Finally, the long and short-term neural memory networks, multilayer perceptrons, and nearest neighbor models are considered the least interpretable models and are therefore classified as the third category. Because the structure of Random Forest, Gradient Lifting algorithm, Ensemble Learning

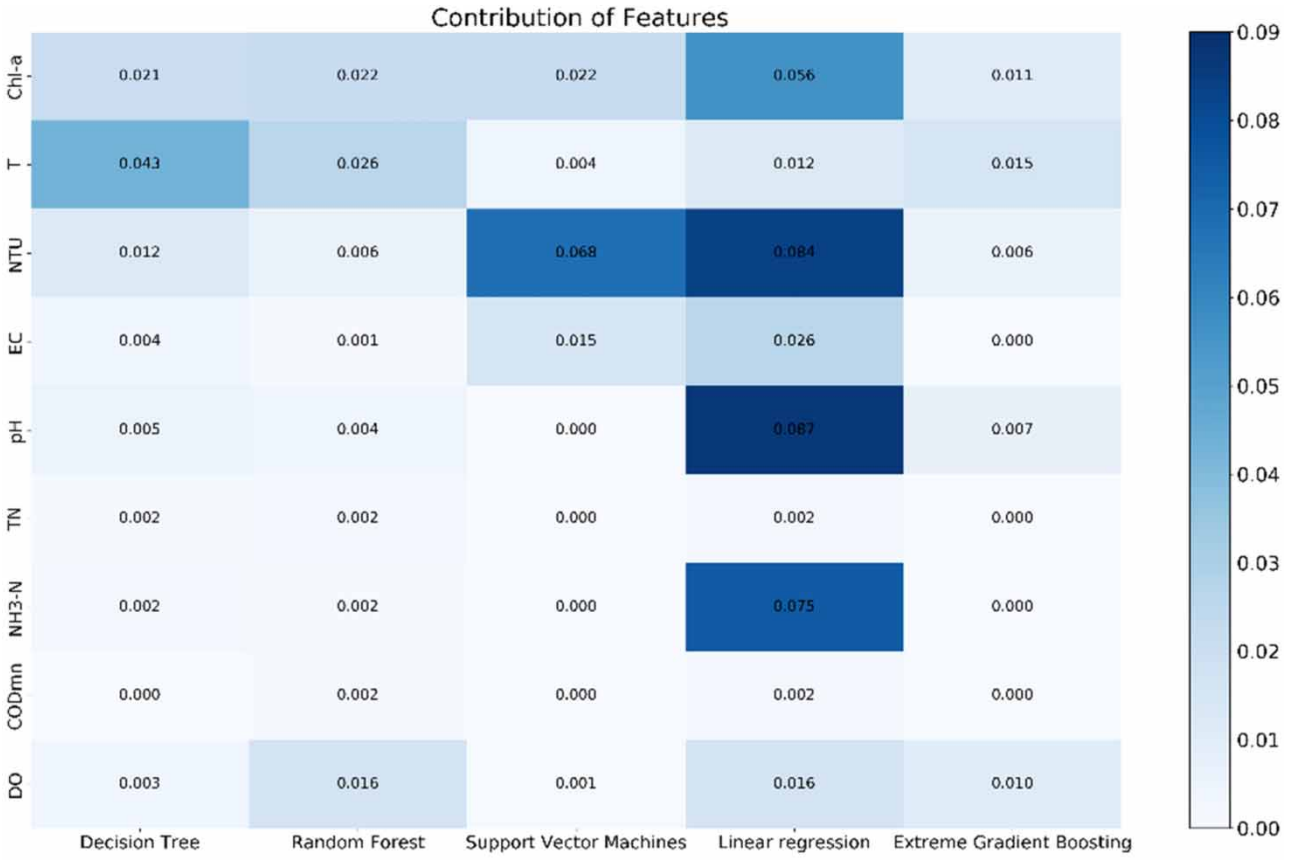


Figure 5 | Correlation between characteristic factors of different models and phosphorus element (the Gehu Lake data).

algorithm and Support Vector Machine is based on Decision Tree, which can be used for visual analysis of the prediction process, they have higher interpretability than LSTM Neural Network, Multilayer Perceptron, and Nearest Neighbor, so they are placed in the middle of the interpretability ranking.

CONCLUSIONS

In this study, we constructed 10 kinds of machine and deep learning models, which all follow the mapping function $f(\theta)$ representation method. Through comprehensive comparison of prediction accuracy and interpretability, a prediction model suitable for different water body types of phosphorus was found. One of the purposes of this study is to apply phosphorus prediction models suitable for different water bodies to government management data platforms, providing prediction and warning of phosphorus in water bodies. Each model is based on high-frequency monitoring data of real water bodies affected by phosphorus and is applied and evaluated. The predictive performance of the model depends on the R^2 value, and the interpretability is evaluated according to the three defined criteria. In addition, the contribution degree of each characteristic factor is quantitatively obtained by using PDP and SHAP methods. The main conclusions of this study are as follows:

1. Among the three evaluation indexes of model interpretability, 10 kinds of model interpretability ranking are obtained. The categories ranked in class I include linear regression, multiple regression and decision tree, and the categories ranked in Class II include Random Forest, Limiting Gradient Random Number, Integrated Learning algorithm and Support Vector Machine. Class III consists of Short- and Long-time Memory Neural Networks, Multilayer Perceptron, and Nearest Neighbor algorithm.
2. The Stochastic Forest, combined linear regression and Stochastic Forest models are suitable for three different water datasets, including the Yangtze River, the Beijing-Hangzhou Canal and Gehu Lake, respectively, from the perspective of the prediction accuracy and comprehensive evaluation of the model's interpretability. Among them, the comprehensive level of nearest neighbor and deep learning models is at a low level, and these models are usually considered as 'winners of competition'. We explained the

Table 4 | Interpretability ranking of different models

Model	Interpretability criteria for total phosphorus	Interpretable evaluation	Explainable ranking
Linear regression	High transparency; Accurate response relationship; Accurate feature importance ranking	Very high	Class I
Multiple regression	High transparency; Accurate response relationship; Accurate feature importance ranking	Very high	
Decision tree	High transparency; Approximate response relationship; Approximate feature importance ranking	High	
Random forest	Low transparency; Approximate response relationship; Approximate feature importance ranking	Medium	Class II
Extreme gradient boosting	Low transparency; Approximate response relationship; Approximate feature importance ranking	Medium	
Integrated approach	Low transparency; Approximate response relationship; Approximate feature importance ranking	Medium	
Support vector machines	Low transparency; Approximate response relationship; Approximate feature importance ranking	Low-middle	Class III
Long short-term memory neural network	Transparency is approximately zero; No response relationship; Approximate feature importance ranking	Low	
Multilayer perceptron	Transparency is approximately zero; Inaccurate response relationship; Approximate feature importance ranking	Low	
Nearest neighbor	Transparency is approximately zero; Inaccurate response relationship; Approximate feature importance ranking	Low	

reason for their lower rankings, as the data used in this article is limited and the deep learning results reported in the literature are based on thousands of training set data. Overall, the experimental results focused on the Gehu dataset in this article indicate that the random forest model undoubtedly performs best based on prediction accuracy and model interpretability.

3. After the combined analysis of the contributions of various features to phosphorus in the Yangtze River, the Beijing-Hangzhou Canal, and Gehu Lake, it shows that water temperature, chlorophyll-a, and turbidity are the three most important factors in the process of model prediction.

The 10 water quality prediction models constructed in this article all come from the fields of machine learning and neural networks. They only conduct interpretability research on predicting phosphorus elements in lakes and rivers at the model data layer, lacking physical and chemical knowledge of lake and river water quality. In the future, we want to build a comprehensive model of water quality mechanism coupled neural network and conduct interpretability research on it, so that the model can better match the actual situation of phosphorus elements in lake and river water quality.

ACKNOWLEDGEMENTS

This study is funded by Changzhou Key Research and Development Plan (Science and Technology Support for Social Development) project (CE20225061), Jiangsu Agriculture Science and Technology Innovation Fund (JASTIF(CX(22)3111), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX23_3067), and Innovation and Entrepreneurship Training Program of Jiangsu College Student (202310292093Y, 202310292078Y).

DATA AVAILABILITY STATEMENT

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

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Barzegar, R., Aalami, M. T. & Adamowski, J. 2020 Short-term water quality variable prediction using a hybrid CNN-LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment* **34** (8), 1–19.

Breiman, L. 2001 Statistical modeling: the two cultures. *Statistical Science* **16**, 199–231.

Chao, B., Can, Y. J., Xu, X. G., Li, M. B. & Hu, G. Z. 2023 Study on lake pollution traceability based on water quality fluorescence fingerprint – a case study of Gehu Lake in the Taihu Lake Lake Basin. *Lake Science* **35** (04), 1330–1342.

Chen, Q., Guan, T., Yun, L., Li, R. & Recknagel, F. 2015 Online forecasting chlorophyll-a concentrations by an auto-regressive integrated moving average model: feasibilities and potentials. *Harmful Algae* **43**, 58–65.

Cheng, Z. H. 2022 *Characteristics and Causes of Natural Water Chemical Changes in the Tao River System of Taihu Lake.*

China ecological encironment bulletin 2022 *Journal of Environmental Protection* **50**, 61–74.

Donoho, D. 2017 50 years of data science. *Journal of Computational and Graphical Statistics* **26**, 745–766.

Doshi-Velez, F. & Kim, B. 2017 Towards a rigorous science of interpretable Machine learning. arXiv 1702.08608.

Elser, J. J. & Patrick, H. R. 1994 A stoichiometric analysis of the zooplank ton-phytoplankton interaction in marine and freshwater ecosystems. *Nature* **370**, 211–213.

Hans, V., Niels, E., Arjan, B., Kumar, D. & Pandey, L. K. 2022 What drives the ecological quality of surface waters? A review of 11 predictive modeling tools. *Water Research* **12** (05), 326–334.

Huang, J., Zhang, Y., Arhonditsis, G. B., Gao, J. & Peng, J. 2020 The magnitude and divers of harmful algal blooms in China's lakes and reservoirs: a national-scale characterization. *Water Research* **181**, 115902.

Koch, G., Kuhni, M., Rieger, L. & Siegrist, H. 2001 Calibration and validation of an ASM3-based steady-state model for activated sludge systems – part II: prediction of phosphorus removal. *Water Research* **35** (9), 2246–2255.

Kuhn, M. & Johnson, K. 2016 *Applied Predictive Modeling*. Springer, New York, NY, USA.

Li, J. D., Li, Y. M., Lv, H., Dong, X. Z., Can, X. L. & Zeng, S. 2022 Vertical distribution characteristics and dynamic mechanism of cyanobacteria in shallow eutrophic lakes: a case study of the Taihu Lake Lake. *Journal of Environmental Science* **42** (07), 318–328.

Liu, Z. 2020 Bidirectional ecological compensation mechanism and accounting method in plain river network area. *Research of Environmental Science* **33** (11), 2554–2560.

Liu, X. M. & Zhang, G. X. 2022 A review on the impact of climate change on blue algae bloom in lakes. *Advance in Water Science* **33** (02), 316–326.

Molnar, C. 2021 *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. Interpretable-ml-book*.

Ouyang, T., Shan, K., Zhou, B. T., Huang, L., Wu, Z. X. & Shang, M. S. 2021 Research on online algae time series data prediction based on LSTM network: taking the three gorges reservoir as an example. *Lake Science* **33** (04), 1031–1042.

Qin, B. Q., Zhou, J. & Elster, J. J. 2020 Water depth underpins the relative roles and fates of nitrogen and phosphorus in lakes. *Environmental Science & Technology* **54** (6), 3191–3198.

Recknagel, F., Orr, P., Bartkow, M., Swanepoel, A. & Cao, H. 2017 Early warning of limit- exceeding concentrations of cyanobacteria and cyanotoxins in drinking water reservoirs by inferential modelling. *Harmful Algae* **69**, 18–27.

Reichwaldt, E. S. & Ghadouani, A. 2012 Effects of rainfall patterns on toxic cyanobac-terial blooms in a changing climate: between simplistic scenarios and complex dynamics. *Water Research* **46** (5), 1372–1393.

Sondergaard, M., Jensen, P. J. & Jeppesen, E. 2001 Retention and internal loading of phosphorus in shallow, eutrophic lakes. *The Scientific World Journal* **1**, 427–442.

Thomas, M. K., Fontana, S., Reyes, M., Kehoe, M. & Pomati, F. 2018 The predictability of a lake phytoplankton community, over time-scales of hours to years. *Ecology Letters* **21** (5), 619–628.

Thomson-Laing, G., Puddick, J. & Wood, S. A. 2020 Predicting cyanobacterial biovolumes from phycocyanin fluorescence using a handheld fluorometer in the field. *Harmful Algae* **97**, 101869.

Tong, Y., Xu, X., Zhang, S., Shi, L., Zhang, X., Wang, M., Qi, M., Chen, C., Wen, Y., Zhao, Y., Zhang, W. & Lu, X. 2019 Establishment of season-specific nutrient thresholds and analyses of the effects of nutrient management in eutrophic lakes through statistical machine learning. *Journal of Hydrology* **578**, 124079.

Wu, Z., Wu, S. F. & Liu, Y. 2018 Key processes and quantitative identification methods of nitrogen and phosphorus cycle in lakes. *Journal of Peking University (Natural Science Edition)* **54** (01), 218–228.

Wu, H. Q., Li, Q. H., Li, Q., Gu, P., Zhen, Z. & Zhang, W. Z. 2022 Study on spatial-temporal changes of water eutrophication in the cyanobacteria concentration area in the north of the Taihu Lake. *Environmental Pollution and Prevention* **44** (07), 926–932.

Xia, R., Wang, G., Zhang, Y., Yang, P., Yang, Z., Ding, S., Jia, X., Yang, C., Liu, C., Ma, S., Lin, J., Wang, X., Hou, X., Zhang, K., Gao, X., Duan, P. & Qian, C. 2020 River algal blooms are well predicted by antecedent environmental conditions. *Water Research* **185**, 116221.

Xu, H., Paerl, H. W. & Qin, B. 2015 Determining critical nutrient thresholds needed to control harmful cyanobacterial blooms in eutrophic lake Taihu, China. *Environment Science & Technology* **49** (2), 1051–1059.

Yang, Y. F., Wu, J., Wang, L., Guo, Y., Hui, T. T., Gao, W. & Zhang, Y. 2022 Simulation and prediction of high time resolution nitrogen and phosphorus concentrations in Liaohe River based on random forest model. *Journal of Environmental Science* **42** (12), 384–391.

Yarkoni, T. & Westfall, J. 2017 Choosing prediction over explanation in psychology: lessons from Machine learning. *Perspect. Psychol. Sci.* **12** (6), 1100–1122.

Zhang, C., Huang W, F. & Li, r. 2022 The effect of salinity on nitrogen and phosphorus nutrient excretion during lake ice freezing. *Lake Science* **34** (04), 1186–1196.

Zhu, M. Y., Zhu, G. W. & Nurminen, L. 2015 The influence of macrophytes on sediment resuspension and the effect of associated nutrients in a shallow and large lake. *PLoS One* **10** (6), e0127915.

First received 20 July 2023; accepted in revised form 13 September 2023. Available online 26 September 2023