Development of Mass Proliferation Control Algorithm of Phytoplankton Using Artificial Neural Network

Seonghwa Park[*] · Jonggu Kim[**] · Minsun Kwon[***†]

* PhD Candidate, Dept. of Civil & Environmental Engineering, Kunsan National University, Kunsan 54150, Korea
** Professor, Dept. of Environmental Engineering, Kunsan National University, Kunsan 54150, Korea
*** PhD, Ocean Physics Dept., Land & Ocean Environmental Eng., Suwon 16690, Korea

인공신경망을 이용한 식물플랑크톤의 대량 증식 제어 알고리즘 개발

박성화[*] · 김종구[**] · 권민선[***†]

* 군산대학교 토목환경공학부 박사과정, ** 군산대학교 환경공학과 교수, *** 국토해양환경기술단 연구원

Abstract : *Suitable environmental conditions in Saemangeum frequently favor phytoplankton growth. There have been occurrences of sudden phytoplankton blooms, surpassing the algae management standards. A model was designed to prevent such blooms using scientific predictive techniques to forecast and regulate the possibility of phytoplankton blooms. We propose effective and efficient algae control measures concerning every phytoplankton species optimized through the policy control of nutrients (DIN, PO4-P) from rivers and controlling lake salinity using gate operations. The probability of phytoplankton blooms was initially forecast using an artificial neural network algorithm based on observations. The model's Kappa number fluctuated from 0.7889 to 1.0000, indicating good to excellent predictive power. The Garson algorithm was then utilized to assess the significance of explanatory variables for every species. Meanwhile, the probability of phytoplankton blooms was anticipated depending on the DIN and salinity value changes. Therefore, the model predicted the precise DIN and salinity concentrations to inhibit phytoplankton blooms for each species. Hence, the green algae model can create effective proactive measures to avoid future phytoplankton blooms in enormous artificial lakes.*

Key Words : *Saemangeum, Phytoplankton, Algal bloom, Artificial neural network, Proliferation control*

요   약 : 새만금 내에서는 종종 식물플랑크톤이 증식하기에 알맞은 환경조건이 생성되며 일시에 식물플랑크톤 대증식이 발생하면서 조류 관리기준을 초과하는 사례가 발생하고 있다. 이를 대비하기 위하여 과학적 예측기법을 토대로, 식물플랑크톤의 종별로 가장 효과적이고 효율적인 녹조발생 억제 방안을 제안하기 위하여 식물플랑크톤 대증식 가능성을 예측하고, 제어할 수 있는 모델을 개발하였다. 즉, 하천에서 유입하는 영양염(DIN, PO4-P)을 정책적으로 조절하고, 갑문운영을 통해 호 내 염분을 제어하는 것이다. 먼저 관측치로부터 인공신경망 알고리즘을 이용해 식물플랑크톤 대증식 가능성을 예측 결과, 모델의 Kappa 수는 0.7889～1.0000의 범위로, good～excellent 수준이었다. 다음으로 Garson 알고리즘을 이용하여 종별로 설명변수의 중요도를 평가하였고, 또한 DIN 및 염분 값의 변화에 따른 식물플랑크톤 대량 증식 확률을 예측하였다. 그 결과, 각 종별로 식물플랑크톤의 대증식을 억제할 수 있는 DIN과 염분 농도를 정량적으로 예측할 수 있었다. 따라서, 향후 새만금과 같은 거대한 인공 호수에서 식물플랑크톤의 대증식을 억제하기 위한 효율적이고 효과적인 대응방안을 마련할 수 있도록 녹조제어모델을 활용할 수 있을 것으로 판단된다.

핵심용어 : 새만금, 식물플랑크톤, 조류대증식, 인공신경망, 녹조제어

* First Author : andriack@kunsan.ac.kr, 063-469-1871
† Corresponding Author : mskwon@landocean.co.kr, 031-695-3474

1. Introtion

The water quality of Saemangeum Lake is deteriorating due to eutrophication, as nutrient-rich freshwater inflows from nearby industrial complexes and the Saemangeum Seawall (33km) was completed in April 2006, and seawater distribution is not sufficient. The distribution of seawater for this deteriorated water quality is carried out through the Sinsi Gate and Garyeok Gate installed on the Saemangeum Seawall, which creates suitable environmental conditions for phytoplankton to grow, and there have been cases of phytoplankton blooms that exceed the algae management standards. Damage caused by green and red tides due to this bloom is a concern.

There are studies on the distribution of phytoplankton in Saemangeum Lake, and Kim et al. (2009) reported changes in phytoplankton communities and distinct seasonal cycles due to semi-diurnal tidal coupling in the lower section of the Mankyeong River before the construction of the Saemangeum Seawall (1999~2000). In addition, Jang et al. (2009) reported a decrease in the number of species and an increase in the abundance of phytoplankton communities compared to previous studies based on surveys at a fixed station near Mankyeong Bridge immediately after the completion of the Saemangeum seawall (2006-2007). Yeo (2010) monitored the biomass of phytoplankton, which is the core of the green and red tide problem in the study area, in terms of abundance (cells/ml) for a long period of time (2001-2010). As a result, the temporal and spatial variability of the study area was examined by dividing the study area into rivers, artificial lakes, and seas. It has been reported that frequent algae blooms occurred in the streams flowing into Saemangeum Lake, and that the planned waters of Mankyeong Lake and Dongjin Lake experienced rapid changes in phytoplankton abundance due to changes in freshwater and seawater inputs and seasonal changes (Yeo, 2012), and although several studies have been reported, there is a lack of research on the prediction and analysis of phytoplankton according to the distribution of seawater.

Various water quality problems, including harmful algal bloom, are occurring worldwide in river-type lakes where sufficient nutrients are supplied at the time when water temperature and light conditions suitable for algae growth are formed. Direct problems caused by the massive proliferation of algae include the toxicity of species such as cyanobacteria (Codd et al., 2005; Lehman et al., 2005), the increased production of volatile organic compounds (VOC) by algae, resulting in bad taste in water supply (Watson, 2004), clogging of filter paper by diatoms (Jun et al., 2001), and human health threats and aesthetic effects due to toxins (Lee et al., 2013; Dencheva, 2010; Li et al., 2011).

In Korea, the algae warning system was piloted in Daecheong Lake in 1996 and expanded to 22 lakes nationwide in 2012 (Lee et al., 2012a), and since 2012, a water quality forecasting system has been implemented for the main stretches of the four major rivers for the purpose of proactive water quality management by predicting short-term changes in water temperature and Chl-a concentration (Lee et al., 2012b). Since May 2020, an algae prediction system has been implemented and operated by integrating the water quality forecasting system and the algae warning system to predict changes in water quality and algae outbreaks in public waters (MOE Order No. 1456).

To prevent this, studies have been reported on phytoplankton prediction in rivers and lakes. Looking at international cases, Recknagel et al. (1994) built an algae bloom prediction model using water quality data observed for 12 years as input to an artificial neural network, Wilson and Recknagel (2001) built an algae bloom prediction model using water quality data observed for 12 years as input to an artificial neural network and conducted model validation, and Karula et al. (2000) built a eutrophication neural network prediction model with a Levenberg-Marquardt (tangent-sigmoid) structure to analyze and predict Chl-a considering various water quality factors. Singh et al. (2009) built a DO and BOD prediction model using BPNN to predict DO and BOD, respectively, for water quality management in rivers.

In Korea, Ahn et al. (2001) performed monthly water quality predictions for DO, BOD, and TN at Gongju Branch of the Geumgang River Basin using the BP algorithmic neural network model and examined its applicability by comparing it with the ARIMA model, and Oh et al. (2002) built an optimal water quality prediction model through monthly water quality predictions for each water quality element using the BP algorithmic neural network model with DO, BOD, TN, and TP data from the Yeongsan River Basin. Lee and Seo (2002) conducted monthly water quality predictions of BOD, TN, and TP concentrations using the WASP5 model to identify the effects on the inflow water quality of Daecheong Lake. Park and Ha (2003) used Genetic Algorithm and Neural Network (GANN) to predict the monthly water quality of DO, BOD, TN, and TP concentrations in the Naju branch of Yeongsan River, and Cho et al. (2004) used BP algorithm neural network model to predict BOD, TN, TP, and TOC concentrations in the Naesacheon and Pyeongchang River

basins within the Chungju Lake basin in real time. Ahn et al. (2000), used the BP algorithm neural network model to build an intelligent monthly water quality prediction model using each water quality data of the Dalcheon branch of the Han River basin and verified its applicability. Oh et al. (2008) developed a daily prediction model for runoff, TOC, and TOC load at the Naju branch of the Yeongsan River basin using the BPNN model. It was also used for the development and application of algae simulation techniques using chlorophyll-a concentration and cell counts by algae species in Lake Uiam in the midstream of the Bukhan River (Choi et al., 2015). However, it is difficult to apply it to Saemangeum, which has the characteristic that seawater is distributed through locks.

Meanwhile, Park et al. (2023) demonstrated using taxonomic statistics that salinity, including phototrophic salinity, is linked to the presence of phytoplankton. Consequently, they deduced that algal blooms' likelihood could be affected by shifts in salinity via the drainage gate.

This study does not aim to quantitatively predict the abundance or biomass of phytoplankton. Rather, it uses a classification approach to predict the probability of algal blooms. Using this approach to derive quantitative amounts of each controlling factor, allows for the calculation of the concentration of salts that can inhibit algal blooms. Algae blooms are significantly influenced by nutrients. With that said, predicting the probability of algae blooms using machine learning algorithms can allow for calculating the concentration of nutrients that can suppress algae blooms. In summary, this study aims to propose the most effective and efficient algae bloom suppression measures for each species of phytoplankton at each point in Saemangeum based on scientific prediction techniques.

2. Material and Method

2.1 Algal bloom control model design

The model is designed for future data accumulation. Data is collected in real-time or intermittently, and the collected data is preprocessed and stored in a data archive. The user selects a target species and a training dataset to predict it. The model predicts the probability of an algae bloom, which is then calibrated based on the model's confidence in the target species. Once you have the confidence, you select the variables you want to control and use algebra to predict the quantitative amount of the variable that will suppress algae growth. The final decision is whether to control or

not, and the variable is controlled based on the result.

In this study, the observed data of 2021 were preprocessed and stored in the data archive, and a training dataset was created to fit the model using an artificial neural network algorithm. The training dataset of the model consists of 2,556 rows with 45 columns, including vertex, observation date, water quality, species abundance, month, temperature, precipitation, insolation, and evaporation. Using this, the model was fitted for each species, and the quantitative value of the target variable that reduces the probability of phytoplankton blooms was predicted by substituting the explanatory variables and the target data (DIN, Salinity).
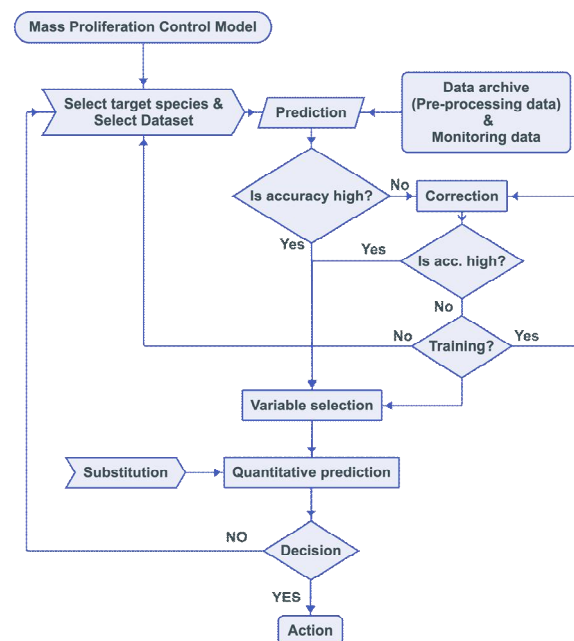

Fig. 1. Flow chart of algae control model.

2.2 Data and preprocessing

The data used in the model were observed once or twice a month from January to November 2021, including 10 months (January 25, February 22, March 24, April 26, May 19, June 29, July 13, August 9, September 8, October 13, and November 1) and survey during summer rainfall (August 29 and September 30). A total of seven observation locations (Fig. 2) were selected based on the water quality measurement network points in Saemangeum Lake, which are being investigated by the Jeonbuk Provincial Environment Agency.

An Ocean Seven 310 CTD from Idronaut (Italy) was used for the observations, and the specifications of the instrument are presented in Table 1.

Table 1. Specifications of CTD observation device

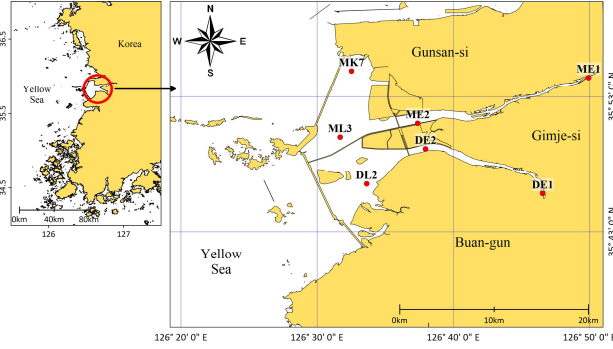| item	measuring range	precision	resolution
Salinity	0~70mS/cm	0.005mS/cm	0.001mS/cm
Temperature	-3~50℃	±0.005℃	±0.0005℃
Pressure	0~1000dbar	0.05%	0.0015%

Fig. 2. Inner and outer observation vertices in Saemangeum.

The target phytoplankton are *Skeletonema* spp., *Cyclotella atomus*, *Stephanodiscus*, *Chaetoceros* spp. *Phormidium tenue*. To compensate for the lack of data prior to model design, a piecewise cubic Hermitian polynomial interpolation was used, which captures the motion of the data well while suppressing exaggerated values as much as possible. If the piecewise cubic polynomial is $P(x)$, then in a two-dimensional coordinate system consisting of (x,y), h_x and δ_x are defined as follows.

$$h_x = x_{k+1} - x_k \tag{1}$$

$$\delta_k = \frac{y_{k+1} - y_k}{h_k} \tag{2}$$

In addition, the slope of $P(x)$ at x_k can be expressed as $d_x = P'(x)$, and if $s = x - x_k, h = h_k$ is in the range of $x_k \leqq x \leqq x_{k+1}$, the cubic equation $P(x)$ can be expressed as follows.

$$P(x) = \frac{3hs^2 - 2s^3}{h^3}y_{k+1} + \frac{h^3 - 3hs^3 + s^3}{h^3}y_k + \frac{s^3(s-h)}{h^2}d_{k+1} + \frac{s(s-h)^3}{h^2}d_k \tag{3}$$

As above, the cubic polynomial $P(x)$ expressed by s and x is called a piecewise cubic Hermitian interpolation polynomial. The above equation requires 4 interpolation conditions, which are represented by 2 function values and 2 derivative values at a specific point as follows.

$$P(x_k) = y_k, \qquad P(x_{k+1}) = y_{k+1} \tag{4}$$

$$P'(x_k) = d_k, \qquad P'(x_{k+1}) = d_{k+1}$$

2.3 Prediction of phytoplankton overgrowth potential
2.3.1 Summary

Since there are limitations in quantitative prediction of algal organisms, efficiency and accuracy can be maximized by simplifying the problem to whether or not algae proliferate. Therefore, the response variable becomes a qualitative or categorical variable as opposed to a continuous or quantitative variable, and in this study, a classification algorithm that predicts qualitative variables among machine learning algorithms was used.

In this study, the Artificial Neural Network algorithm was adopted, but since there is not much data accumulated so far, we focused on the design of the model without distinguishing between training data and target data. On the other hand, an artificial neural network is an algorithm for machine learning, that is, machine learning developed inspired by human nerves. In general, a multilayer artificial neural network is divided into three layers: an input layer, a hidden layer, and an output layer, and each layer is composed of nodes. The input layer is composed of supply neurons and serves to input the values of predictor variables for deriving a predicted value. If there are n input values, the input layer has n nodes. The hidden layer consists of computational neurons, receives input values from input nodes, calculates a weighted sum, applies this value to a transition function, and delivers it to the output layer. When an input signal x is received and y is output, it can be expressed as $y = wx + b$, where w is a weight and b is a bias. In other words, a general artificial neuron with n number of input protrusions is expressed as follows.

$$y(x) = f\left(\sum_{i=1}^{n} w_i x_i\right) \tag{5}$$

An artificial neural network uses an activation function as a function that converts the sum of input signals into an output signal, and in this study, ReLU (Rectified Linear Unit) function, which is mainly used recently, was used.

$$h(x) = \begin{cases} x \ (x > 0) \\ 0 \ (x \leqq 0) \end{cases} \tag{6}$$

In this study, the number of hidden layers was set to 20, and the weights were initialized to 0 for consistency in prediction.

2.3.2 Determination of explanatory and response variables

Phytoplankton can proliferate under the influence of physical factors such as water temperature and salinity, chemical factors such as nutrients and trace elements, and biological factors such as symbiosis and predation pressure (Kim et al., 2018). Therefore, water temperature, salinity, and nutrients (DIN, DIP) are the most basic factors to be considered.

Since insolation affects the photosynthesis of phytoplankton and rainfall determines the transport of nutrients in lakes such as Saemangeum, these two meteorological factors were included. On the other hand, as an important matter to be considered for the control of algal bloom, real-time monitoring or equivalent quick and simple observation should be possible, so biological factors were excluded. Therefore, as explanatory variables, environmental factors such as water temperature and salinity, nutrients of DIN and PO4-P, and meteorological conditions such as insolation and rainfall were determined. In the case of rainfall, the sum of the previous 24 hours based on the observation date was used.

The response variable is a categorical type, and the simpler the category, the higher the efficiency of the model and the higher the prediction accuracy, so it was simplified to Normal and Caution. Caution was determined when the current amount of algae was 1,000 cells/mL or more. The predicted targets are *Skeletonema* spp., *Cyclotella atomus, Stephanodiscus, Chaetoceros* spp., and *Phormidium tenue.*

2.3.3 Performance indicators of the model

As shown in Table 2, when each cell of the confusion matrix is defined as a, b, c, and d, the definition of each performance indicator is as follows.

$$Accuracy = \frac{(a+d)}{(a+b+c+d)} \qquad (7)$$

$$Kappa = \frac{\frac{(a+d)}{(a+b+c+d)} - \frac{(a+b)(a+c)+(c+d)(b+d)}{(a+b+c+d)^2}}{1 - \frac{(a+b)(a+c)+(c+d)(b+d)}{(a+b+c+d)^2}} \qquad (8)$$

$$N.I.R. = \frac{(b+d)}{(a+b+c+d)} \qquad (9)$$

$$Sensitivity = \frac{a}{a+c} \qquad (10)$$

$$Specificity = \frac{d}{b+d} \qquad (11)$$

$$Balanced Accuracy = \frac{Sensitivity + Specificity}{2} \qquad (12)$$

Table 2. Confusion Matrix Structure

Predicticted condition	True condition	
	Positive	Negative
Positive	a : True positive	b : False positive
Negative	c : False negative	d : True negative

Kappa is a statistical metric that measures the agreement between actuals and predictions, with a value of 0 indicating complete disagreement and a value of 1 indicating perfect agreement. The intuitive meaning of the Kappa coefficient is the probability that both the actual value and the observed value match by chance, and a common interpretation of the Kappa coefficient is as shown in Table 3. On the other hand, Balanced accuracy is the average of Sensitivity, the percentage of positive predictions, and Specificity, the percentage of negative predictions. Also, N.I.R. (No Information Rate) is the accuracy when the model predicts only negatives, so Accuracy should be higher than N.I.R..

Table 3. Interpretation criteria for general Kappa statistics

Kappa	Interpretation
0.2 >	Poor agreement
0.2 ~ 0.4	Fair agreement
0.4 ~ 0.6	Moderate agreement
0.6 ~ 0.8	Good agreement
0.8 ~ 1.0	Excellent agreement

2.4 Variable importance

The model first identifies the importance of each explanatory variable, and uses a method of continuously calculating the probability of algal bloom by linearly increasing or decreasing the values of variables with higher importance. The importance of the explanatory variable can be identified as the connection strength between the input node and the hidden node using the Garson algorithm (Garson, 1991). If the input is 'I', the output is 'o', and the relative importance is R, Garson's algorithm is as follows.

$$R = \frac{\sum_{j=1}^{nh} \frac{w_{ji}}{\sum_{l=1}^{\ni} w_{jl}} \cdot w_{oj}}{\sum_{k=1}^{\ni} (\sum_{j=1}^{nh} \frac{w_{jk}}{\sum_{l=1}^{\ni} w_{jl}} \cdot w_{oj})} \qquad (13)$$

Here, ni is the number of input nodes, nh is the number of hidden nodes, and no is the number of output nodes. w_{jl} is the weight between the input node i and the hidden node j, and w_{oj} is the weight between the hidden node j and the output node o.

Using this, the relative importance of explanatory variables is identified, and controllable factors are used. In this model, DIN and salinity were used as control factors.

2.5 Initial conditions for species-specific prediction.

Based on the observations, the conditions of the observation day when the predicted value was predicted as a caveat among the values with a large existing amount were set as initial conditions as shown in Table 4, and DIN and salinity were increased and decreased according to the direction of increase and decrease (dir.) at regular intervals (int.) from minimum (min.) to maximum (max.) as shown in Table 5 to predict the possibility of phytoplankton bloom.

Table 4. Initial conditions for species-specific prediction

	Skele.	Cyclo.	Steph.	Chaet.	Phorm.
Station	ML3	ME1	ME1	ML3	DE1
Temp.(℃)	27.666	18.424	8.200	4.057	16.659
Salinity(ppt)	9.839	1.879	10.500	24.110	0.132
DIN(mg/L)	0.634	4.149	10.164	0.626	3.129
PO4(mg/L)	0.033	0.025	0.010	0.004	0.005
Aboundance (cells/mL)	26,094	14,313	28,800	1,102	8,539
rain_1d (mm/hr)	5.50	0.00	0.00	2.30	0.00
total_rad (MJ/m²)	11.16	25.81	14.70	5.44	0.00
P	0.63	1.00	1.00	1.00	1.00
Level	caution	caution	caution	caution	caution

Table 5. Increment and decrement conditions for variables

		Skele.	Cyclo.	Steph.	Chaet.	Phorm.
DIN (mg/L)	max.	0.634	4.149	10.164	0.626	3.129
	min.	0.046	0.082	0.364	0.136	0.042
	int.	-0.012	-0.083	-0.200	-0.010	-0.063
	dir.	decrease	decrease	decrease	decrease	decrease
Salinity (ppt)	max.	9.839	6.779	20.300	24.110	9.932
	min.	0.039	1.879	10.500	14.310	0.132
	int.	-0.200	+0.100	+0.200	-0.200	+0.200
	dir.	decrease	increase	increase	decrease	increase

3. Results and Analysis

3.1 Artificial Neural Network Algorithm Fit Result

Fig. 3 shows the fitting result of this neural network model, which consists of 6 input nodes, 20 hidden nodes, and 1 output node. Each input node, hidden node, and output node is connected to a network with a weight, which is expressed as the connection strength. In this study, the ReLU (Rectified Linear Unit) function was used as the activation function to convert the sum of input signals into output signals.

Meanwhile, Fig. 4 is the confusion matrix showing how well the fitted model predicted caution and normal. Table 6 shows the performance metrics of the model calculated based on the confusion matrix. The balanced accuracy of the fitted model is 0.9014, 0.8980, 1.0000, 1.0000, and 0.9330 for *Skeletonema* spp., *Cyclotella atomus*, *Stephanodiscus*, *Chaetoceros* spp. and *Phormidium tenue*, respectively. In addition, Kappa values ranged from 0.7889 to 1.0000, indicating good or excellent agreement.
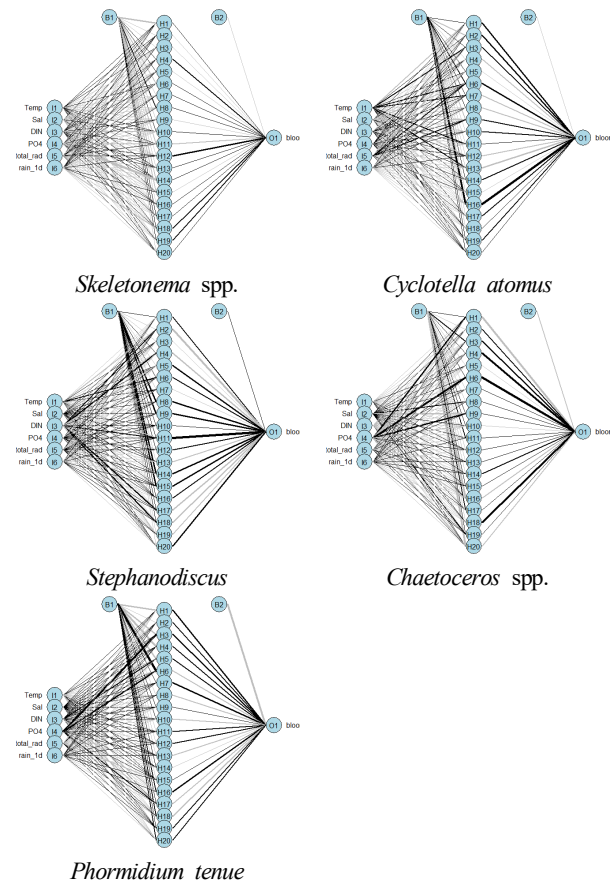

Skeletonema spp.   *Cyclotella atomus*

Stephanodiscus   *Chaetoceros* spp.

Phormidium tenue

Fig. 3. Artificial Neural Network Fitting Result.
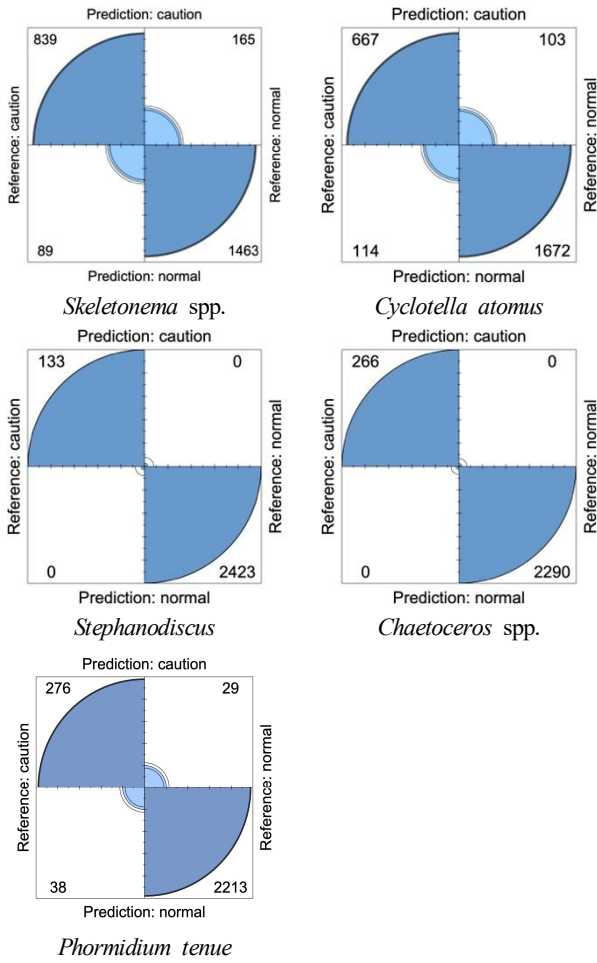
Fig. 4. Confusion matrix of the fitted model.

Table 6. Model's performance metrics

Indicators	*Skele.*	*Cyclo.*	*Steph.*	*Chaet.*	*Phorm.*
Accuracy	0.9006	0.9151	1.0000	1.0000	0.9738
N.I.R.	0.6369	0.6944	0.9480	0.8959	0.8772
Sensitivity	0.9041	0.8540	1.0000	1.0000	0.8790
Specificity	0.8986	0.9420	1.0000	1.0000	0.9871
Balanced Accuracy	0.9014	0.8980	1.0000	1.0000	0.9330
Kappa	0.7889	0.7992	1.0000	1.0000	0.8769
	Good	Good	Excellent	Excellent	Excellent

3.2 Importance of explanatory variables

The Garson algorithm was used to determine the importance of each explanatory variable. The results are shown in Fig. 5.

3.2.1 *Skeletonema* spp.

Looking at the initial conditions in Table 4, when *Skeletonema*

spp. The importance of variables was in the order of PO4 > Salinity > DIN > Solar Radiation > Water Temperature > Rainfall. On the other hand, as shown in Table 1, when all species proliferated in large quantities, PO4-P was at a very low level, so controlling it is meaningless. Therefore, the mass growth probability according to DIN and salt concentration was calculated.

On the other hand, according to the study of Park et al. (2023), the mass growth of *Skeletonema* spp. is suppressed when there is no influx of salt, and it can be interpreted as mass growth when salt is introduced, so it was changed in the direction of reducing salinity.

3.2.2 *Cyclotella atomus*

When *Cyclotella atomus* proliferated in large quantities, the salinity was about 1.879 ppt, which was close to that of fresh water, and DIN was 4.149 mg/L and PO4-P was 0.025 mg/L. The importance of variables appeared in the order of PO4 > DIN > Salinity > Water Temperature > Insolation > Rainfall. Salinity was changed in the direction of increasing, and since DIN is very high, it is considered that the effect of inhibiting the growth of phytoplankton can be increased by limiting the inflow.

3.2.3 *Stephanodiscus*

At the time of the *Stephanodiscus* bloom, salinity was around 10.5000 ppt, brackish water conditions, DIN was 10.1640 mg/L, and PO4-P was 0.0100 mg/L. The order of importance of the variables was Water Temperature > DIN > Salinity > Insolation > PO4 > Rainfall. The salinity was changed in the direction of increasing, and since the concentration of DIN is very high, it is judged that the effect of inhibiting the proliferation of phytoplankton can be improved by limiting it.

3.2.4 *Cheatoceros* spp.

At the time *Cheatoceros* spp. proliferated in large quantities, salinity was about 24.110 ppt in brackish water conditions, DIN was 0.626 mg/L, and PO4-P was 0.004 mg/L. The importance of variables was in the order of DIN > Water Temperature > Rainfall > Salinity > Solar Radiation > PO4. Salinity was changed in a decreasing direction.

3.2.5 *Phormidium tenue*

At the time of the *Phormidium tenue* bloom, salinity was about 0.1319 ppt, which is freshwater conditions, DIN was 3.1288 mg/L, and PO4-P was 0.0050 mg/L. The order of importance of the variables was PO4 > DIN > Salinity > Water Temperature >

Insolation > Rainfall. Salinity is increased by opening the gate, and DIN is very high, so restricting the inflow will inhibit phytoplankton growth.
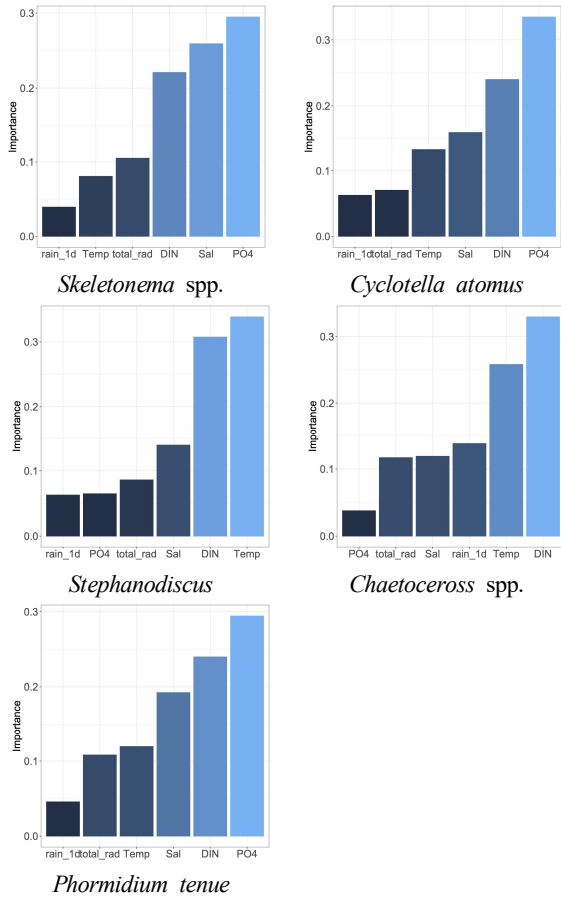

Fig. 5. Importance of each input node in artificial neural network model.

3.3 Algal bloom control model prediction result

For each species, we quantitatively predicted the level of salinity and DIN that should be maintained to inhibit phytoplankton blooms under randomized conditions where blooms occurred (Fig. 6). PO4 was excluded from the calculation because it is present at too low a level (less than 0.1 mg/L), even though it is important, while nutrients were calculated as the probability of increasing or decreasing DIN.

3.3.1 *Skeletonema* spp.

For *Skeletonema* spp. the probability of mass proliferation decreased from about 63.3% to dir 49.9% when DIN was lowered from 0.634 mg/L to 0.130 mg/L. Mass proliferation was predicted to be inhibited when salinity was between 6.039 and 8.439 ppt and below 1.839 ppt.

3.3.2 *Cyclotella atomus*

For *Cyclotella atomus*, lowering DIN from 4.149 mg/L to 0.165 mg/L reduced the probability of mass proliferation from about 100.0% to 7.8%. Mass proliferation was predicted to be inhibited at salinities above about 4.379 ppt.

3.3.3 *Stephanodiscus*

For *Stephanodiscus*, lowering DIN from 10.164 mg/L to 8.364 mg/L reduced the probability of mass proliferation from about 100.0% to 50.0%. Mass growth was predicted to be inhibited at salinities above about 19.5 ppt.

3.3.4 *Chaetoceros* spp.

For *Chaetoceros* spp. the probability of mass proliferation decreased from about 99.9% to 40.2% when DIN was lowered from 0.626 mg/L to 0.296 mg/L. Mass proliferation was predicted to be inhibited at salinities below about 22.310 ppt.

3.3.5 *Phormidium tenue*

For *Phormidium teunue*, lowering DIN from 3.129 mg/L to 0.420 mg/L reduced the probability of mass growth from approximately 100.0% to 0.0%. Mass proliferation was predicted to be inhibited at salinities above about 2.932 ppt.
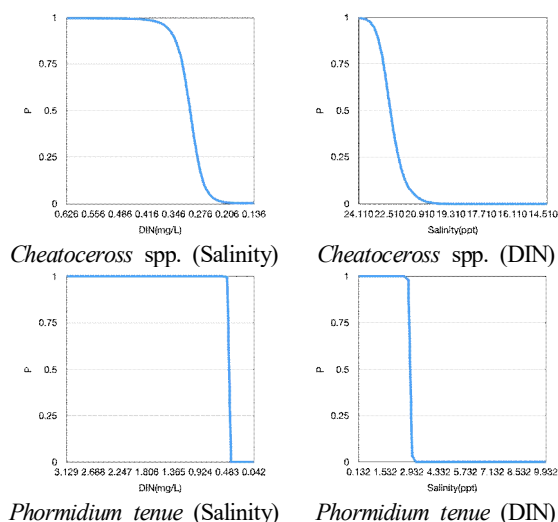
Cheatoceross spp. (Salinity)   *Cheatoceross* spp. (DIN)

Phormidium tenue (Salinity)   *Phormidium tenue* (DIN)
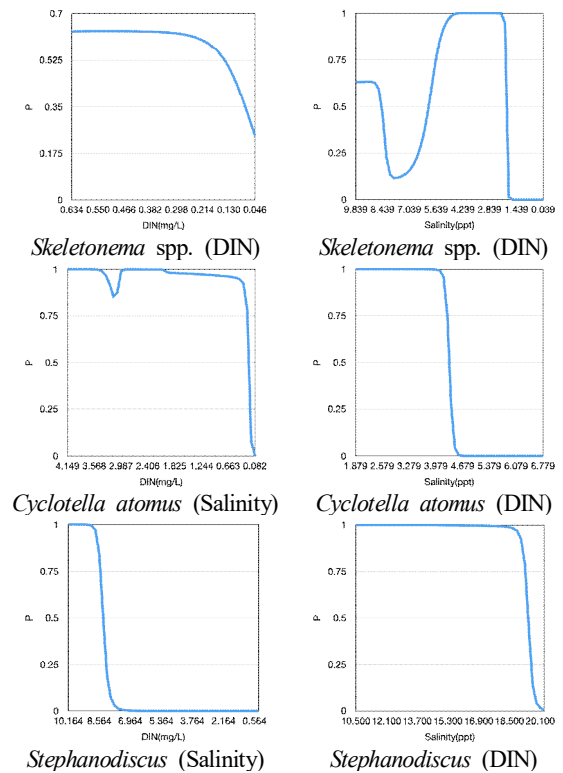
Fig. 6. Probability of phytoplankton blooms as a function of salinity and nutrient concentration.

4. Conclusion

Using an artificial neural network algorithm, we were able to predict the probability of blooms according to phytoplankton species, and predict the quantitative amount of DIN and salinity to suppress blooms, so we were able to prepare efficient and effective countermeasures to control phytoplankton blooms. However, the reliability of the model was not sufficient with only one year of observations, and it will be possible to build a more sophisticated model if additional data can be accumulated in the future. The phytoplankton bloom control model is expected to contribute to the prediction and warning of phytoplankton blooms in large artificial lakes such as Saemangeum, and to efficiently suppress them.

References

[1] Ahn, S. J., I. S. Yeon, Y. S. Han, and J. K. Lee(2001), Water quality forecasting at Gongju station in Geum River using neural network model., Journal of the Korean Society of Civil Engineers, Vol. 34, No. 5, pp. 701-711.

[2] Ahn, S. J., K. W. Jun, and K. I. Kim(2000), Forecastion of Runoff Hydrograph Usiong Neural Network Algorithms, Journal of Korea Water Resources Association, Vol. 33, No. 4.

[3] Cho, Y. J., I. S. Yeon, and J. K. Lee(2004), Application of neural network model to the real-time forecasting of water quality, Journal of Korean Society on Water Quality, Vol. 18, No. 4, pp. 321-326.

[4] Choi, J. K., J. H. Min, and D. W. Kim(2015), Three-dimensional Algal Dynamics Modeling Study in Lake Euiam Based on Limited Monitoring Data. Journal of Korean Society on Water Environment, 31(2), pp. 181-195.

[5] Codd, G. A., L. F. Morrison, and J. S. Metcalf(2005), Cyanobacterial Toxins: Risk Management for Health Protection, Toxicology and Applied Pharmacology, 203, pp. 264-272.

[6] Dencheva, K.(2010), State of Macrophytobenthic Communities and Ecological Status of the Varna Bay, Varna Lakes and Burgas Bay, Phytologia Balcanica, 16(1), pp. 43-50.

[7] Garson, G. D.(1991), Interpreting Neural Network Connection Weights. AI Expert, 6, pp. 47-51.

[8] Jang, K. G., J. W. Park, J. H. Park, N. Ha, and W. H. Yih(2009), Drastic Change of Phytoplankton Community at the Station "Mankyeong Bridge" of the New Saemankeum Lake during 2006-2007. Ocean and Polar Research, 31(1), pp. 71-76.

[9] Jun, H. B., Y. J. Lee, B. D. Lee, and C. J. An(2001), Effects of the Ratio of Diatoms Length to the Effective Size of Filter Medium on Filter Clogging, Journal of the Korean Geo-Environmental Society, 2(1), pp. 31-35. [Korean Literature]

[10] Karula, C., S. Soyupaka, A. F. Cilesizc, N. Akbayb, and E. Germenb(2000), Case studies on the use of neural networks in eutrophication modeling., J. Ecological Modelling, Vol. 134, pp. 145-152.

[11] Kim, H. J., J. Y. Park, and C. H. Moon(2018), Phytoplankton Spring Bloom and Environmental Factors in the Southern East Sea, Korea. THE JOURNAL OF FISHERIES AND MARINE SCIENCES EDUCATION, 30(1), pp. 19-27.

[12] Kim, K. T., E. S. Kim, S. S. Kim, J. S. Park, J. K. Park, and S. R. Cho(2009), Water Quality and Heavy Metals in the Surface Seawaters of the Saemangeum Area during the Saemangeum-dike Construction. Journal of the Korean Society for Marine Environment & Energy, 12(1), pp. 35-46.

[13] Lee, C. S., C. Y. Ahn, H. J. La, S. H. Lee, and H. M. Oh(2013), Technical and Strategic Approach for the Control of Cyanobacterial Bloom in Fresh Waters, Korean Journal of Environment Biology, 31(4), pp. 233-242.

[14] Lee, E. H. and D. I. Seo(2002), Water Quality Modelling of the Keum River - Effect of Yongdam Dam, Journal of Korea Water Resources Association, 35(5), pp. 525-539.

[15] Lee, E. J., E. H. Na, and K. H. Kim(2012a), Development of a Water Quality Forecasting System for a Preventive Water Quality Management, Rural Resources, 54(1), pp. 50-55.

[16] Lee, J. J., I. C. Choi, J. H. Yoon, S. H. Hong, S. Y. Yang, and Y. J. Lee(2012b), Report of Phytoplankton Alert System in the Daechung and Boryung Reservoirs, 11-1480523-001495-01, Guem River Environment Research Center, pp. 1-84. [Korean Literature]

[17] Lehman, P. W., G. Boyer, C. Hall, S. Waller, and K. Gehrts (2005), Distribution and Toxicity of a New Colonial Microcystis aeruginosa Bloom in the San Francisco Bay Estuary, California, Hydrobiologia, 541, pp. 87-99.

[18] Li, J., F. L. Peng, D. B. Ding, S. B. Zhang, D. L. Li, and T. Zhang(2011), Characteristics of the Phytoplankton Community and Bioaccumulation of Heavy Metals during Algal Blooms in Xiangjiang River, Science China Life Sciences, 54, pp. 931-938.

[19] Oh, C. R., S. C. Park, H. M. Lee, and Y. P. Pyo(2002), A forecasting of water quality in the Youngsan River using neural network., Journal of the Korean Society of Civil Engineers, Vol. 35, No. 5, pp. 525-539.

[20] Oh, C. R., Y. H. Jin, D. R. Kim, and S. C. Park(2008), Study on development of artificial neural network forecasting model using runoff, water quality data., Journal of Korea Water Resources Association, Vol. 41, No. 10, pp. 1035-1044.

[21] Park, S. C. and S. J. Ha(2003), Forecasting the water quality of river using GANN., Journal of the Korean Society of Civil Engineers, Vol. 23, No. 6B, pp. 507-514.

[22] Park, S. H., J. G. Kim, G. O. Myung, and M. S. Kwon(2023), Relationship between phytoplankton distribution and salinity in the giant artificial brackish Lake Saemangeum, Ecohydrology & Hydrobiology, Vol. 23, pp. 251-260.

[23] Recknagel, F., M. French, P. Harkonen, and K. I. Yabunaka (1994), Artificial neural network approach for modeling and prediction of algal blooms., J. Ecological Modelling, Vol. 96, pp. 11-28.

[24] Singh, K. P., A. Basant, A. Malik, and G. Jain(2009), Artificial neural network modeling of the river water quality-a case study, J. Ecological Modelling, Vol. 220, pp. 888-895.

[25] Watson, S. B.(2004), Aquatic Taste and Odor: A Primary Signal of Drinking-Water Integrity, Journal of Toxicology and Environmental Health, Part A 67, pp. 1779-1795.

[26] Wilson, H. and F. Recknagel(2001), Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes., J. Ecological Modelling, Vol. 146, pp. 69-84.

[27] Yeo, H. G.(2010), Diversity of planktonic Micro Algae in Saemangeum Water Regions, Journal of the Korea Academia-Industrial Cooperation Society, Vol. 11, No. 9, pp. 3610-3614.

[28] Yeo, H. G.(2012), Annual Variations (2001-2010) of Phytoplankton Standing Stocks in Saemangeum Water Region, Journal of the Korea Academia-Industrial cooperation Society, 13(9), pp. 4326-4333.