International Journal of GEOMATE, Feb., 2025 Vol.28, Issue 126, pp.100-107
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2025.126.g14109
Geotechnique, Construction Materials and Environment

AN ARTIFICIAL INTELLIGENCE MODEL FOR IDENTIFYING RIVER ENVIRONMENT IMPROVEMENT MEASURES

*Shengping Zhang[1] and Jie Qi[2]

[1]Faculty of Urban Science, Meijo University, Japan; [2]School of International Studies, Utsunomiya University, Japan

*Corresponding Author, Received: 28 Nov. 2024, Revised: 12 Dec. 2024, Accepted: 14 Dec. 2024

ABSTRACT: The objective of this study is to identify the most effective measures to improve the environments of Japan's 109 primary watersheds, designated as Class-A Watersheds, which are overseen by the national government. Traditionally, the selection of environmental improvement measures for each watershed has relied heavily on simulations from analytical models and the personal insights of river management officers or environmental planners familiar with specific watersheds. However, these analytical models are constrained by their limited capacity to incorporate a finite array of quantifiable factors, and river management officers or environmental planners, despite their profound expertise in particular rivers, may not possess extensive experience with all significant rivers within a region. This study first designed an artificial intelligence model to evaluate river environments, using a comprehensive dataset that includes both numerical and categorical data for the 109 Class-A watersheds. Furthermore, the reliability of the artificial intelligence model has been verified and applied to identify the most effective environmental improvement measures for each watershed. The study concluded by indicating that the artificial intelligence model is both reliable and useful in determining which improvements will effectively enhance the river environments. This study is anticipated to contribute to the establishment of a more robust and reliable methodology for river environment planning and management.

Keywords: River environment planning, Artificial intelligence model, Model verification, Environment evaluation

1. INTRODUCTION

Despite numerous efforts to improve the river environment in Japan over the last three decades, progress has stagnated during recent years, which has been illustrated clearly in Fig.1 by the unchanged annual achievement rate of the river environment standards for the water quality indicator BOD (Biochemical Oxygen Demand) [1]. While a relatively high achievement rate has been reached overall, further improvements are strongly expected around the country by local residents.

Extensive research and analysis have been conducted on the reasons why significant improvements in the river environment could not be achieved. According to the survey and analysis by the Ministry of Land, Infrastructure, Transport and Tourism of Japan, which manages the country's major rivers [1], the main causes identified were the progression of urbanization and the delay in advanced treatment of domestic wastewater. These have been long-standing issues, and in response, river environment improvement measures such as urban hydrological environment enhancement and promotion of advanced domestic wastewater treatment have been implemented for individual rivers. However, the effects of these measures have not led to substantial improvements in the river environment, as mentioned earlier.

In this study, instead of directly exploring necessary measures for individual rivers, we aim to identify structural and systematic issues in the management of Japan's river environments. Specifically, we will train an artificial intelligence model using water quality environmental information from all major rivers and then use the trained AI model to explore the most effective river environment measures for both all the major rivers as a whole set and individual rivers.

Traditionally, the selection of environmental improvement measures for each watershed has relied heavily on simulations from analytical models and the personal insights of river management officers or environmental planners familiar with specific watersheds. However, these analytical models are constrained by their limited capacity to incorporate a finite array of quantifiable factors, and river management officers or environmental planners, despite their profound expertise in particular rivers, may not possess extensive experience with all significant rivers within a region. This study develops an Artificial Intelligence model to evaluate river environments. The artificial intelligence model has been widely applied to water environment evaluation and planning [2,3]. A well-known reason why an AI model has been chosen is its powerful capability to process various types of water environment data, such as numerical data, categorical data as well as image data, without any specific attention required [4,5]. The artificial intelligence model has also been proven to be a powerful model for exploring the hidden connections between inputs and outputs.

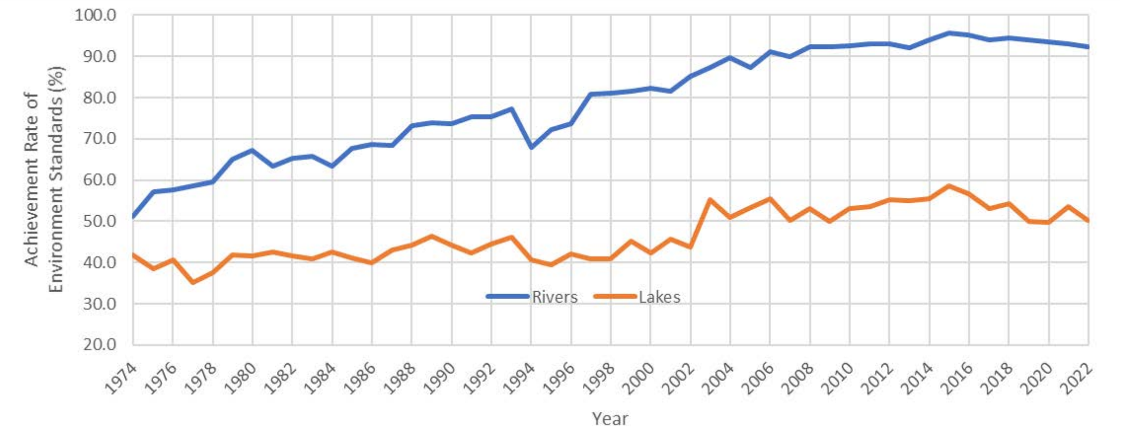

Fig.1 Annual achievement rate of the River Environment Standards for BOD [1]

2. RESEARCH SIGNIFICANCE

This study utilized an artificial intelligence model trained on water environmental data observed across 109 major rivers in Japan to identify the most effective river environmental improvement measures for each river. The term 'effective measures' refers to the environment items that can significantly improve the river environmental ranking across the 109 rivers. To determine the river environmental ranking, an evaluation score model was constructed based on Multivariate Distribution Theory.

3. ARTIFICIAL INTELLIGENCE MODEL

3.1 Structure of a Neural Network

An artificial intelligence model, specifically a neural network model, has been adopted to compose a water environment evaluation method for evaluation or prediction problems due to the suitability of neural network models [4,5].

A neural network is a network system constructed artificially by idealizing the neurons (nerve cells), and consists of a number of nodes and lines that are called *units* and *connections* (or *links*) respectively. Based on the differences in network structures, neural networks generally are classified into two types: layered networks and interconnected networks. It has been shown that a layered network is suitable for evaluation/prediction problems due to its abilities in learning (self-organization) and parallel processing of information.

A typical layered neural network shown in Fig.2, which has a layer of input units at the top, a layer of output units at the bottom, and numerous hidden layers between the input layer and the output layer. Connections exist only between the units in the adjacent layers, and connections within a layer or from higher to lower layers are forbidden.

Theoretically, the neural network model is able to approximate any non-linear relationship between inputs and outputs with any degree of accuracy by using enough hidden layer units and setting connection weights and thresholds to be appropriate through proper learning processes. The potential of this model has been verified with similar problem to this study [3,4].

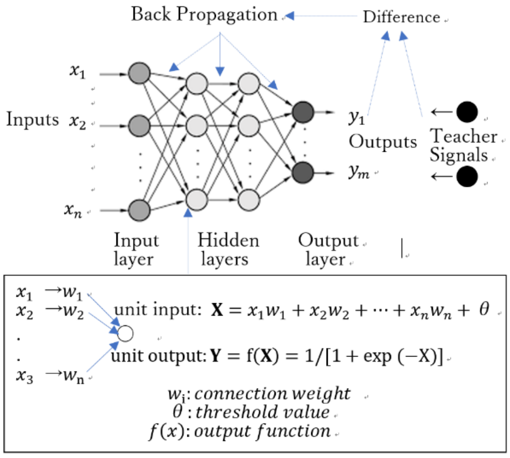

Fig. 2 Structure of a layered neural network.

3.2 Learning Process of Neural Network Model

For a neural network model, the process of setting the connection weights and the unit thresholds is called *learning*. The term *learning* here means the self-organization process through which the neural network model automatically adjusts all the parameters (i.e. all the connections and thresholds) to the appropriate values, when a series of samples of input-output data (called teacher data or teacher signals) are shown to the model. If we consider the information processing in a neural network model as a transformation of input data to output data, then model learning can be considered to be a process through which the neural network model gradually becomes capable of imitating the transforming patterns represented by the teacher data.

Several learning algorithms have been proposed, and among them the Error Back Propagation Algorithm is the most widely used and most successful algorithm [6,7].

To avoid the overfitting problem, a criterion is usually required to make a judgement when the iterative learning process should be terminated. In this study the learning process will be stopped when the Mean Relative Error of the outputs is less than a specified relative error expectation for prediction or evaluation results, which is a common treatment for a learning process of teacher data with random errors (i.e. white noise). Needless to say, this error expectation should be set according to the required accuracy of the problem which is being dealt with.

4. TRAINING NEURAL NETWORK MODEL

4.1 Class-A Rivers and Teacher Data

Japan has 109 important rivers, which have a very large impact on various fields such as flood control, water use, and environmental conservation, and require careful management. Therefore, all these rivers are under the jurisdiction of the central government's Minister of Land, Infrastructure, Transport and Tourism, and are called Class-A rivers.

In this study, the data obtained from the water quality survey conducted for the 109 Class-A rivers of Japan are used for the deep learning process. The data are stored in an open source database that is maintained by Ministry of Land, Infrastructure, Transport and Tourism of Japan [8].

After a careful data verification process, only 104 rivers out of 109 Class-A rivers are chosen to be included in the teacher data set for deep learning because there are quite a few of data missing for the other 5 rivers. For each river the data includes 58 water environment items as shown in Table 1. The data records used in this study are from 1998 to 2018 with a duration of 21 years long.

The 58 environment items are divided into two parts to form a teacher data set, ***evaluation goal variables*** and ***explanation variables***. The evaluation goal variables include the 5 environment items that are used to define The Water Environment Standards for Rivers as shown in Table 2 [9], which are pH, BOD, SS, DO and Total Coliform (TC). All the environment items are used to explain how the achievement of water environment standards are impacted.

4.2 Integrating Evaluation Goal Variables

While it is possible to construct individual artificial intelligence models for each of the five evaluation goal variables adopted in the environmental standards, this approach allows for the identification of effective water environment improvement measures for each evaluation goal variable. However, it does not permit the assessment of how much these measures contribute to the overall improvement of the water environment. Therefore, this study has decided to evaluate the river environment using a single comprehensive evaluation index that integrates the five evaluation goal variables. This comprehensive evaluation index employs a river water quality ranking. A score model for ranking the river environment was constructed as follows.

A score model for ranking the river environment comprehensively has been constructed based on the Multivariate Distribution Theory. Multivariate distributions show comparisons between two or more measurements and the relationships among them. For each univariate distribution with one random variable, there is a more general multivariate distribution. For example, the normal distribution is univariate and its more general counterpart is the multivariate normal distribution, while the multivariate normal model is the most commonly used model for analyzing multivariate data. If the variables are continuous, the multivariate distribution of an N-dimensional variable set $\boldsymbol{x} = (x_1, \cdots, x_N)^{\boldsymbol{T}}$ is described by a density function $f(x_1, \cdots, x_N)$ which satisfied [10]

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, \cdots, x_N) dx_1 \cdots dx_N = 1 \tag{1}$$

There are a variety of multivariate distributions that are well-studied and severed properly for density function $f(x_1, \cdots, x_N)$ such as gamma distribution, logarithmic distribution, and normal distribution. A fitting process is required usually to decide which distribution serves a specific data set the best. However, a fitting process could be very complicated and is usually skipped by directly choosing the normal distribution, which is partially justified by the famous Central Limit Theorem [11].

A Multivariate Normal Distribution of an N-dimensional variable set $\boldsymbol{x} = (x_1; \cdots; x_N)^{\boldsymbol{T}}$ is expressed as follows [11].

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \textstyle\sum) = \frac{1}{\sqrt{(2\pi)^N \det(\sum)}} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^T \sum^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}{2}\right) \tag{2}$$

Where $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \sum)$ is an N-dimensional distribution density function, vector $\boldsymbol{\mu}$ is the mean of the variable set $\boldsymbol{x} = (\boldsymbol{x_1}; \cdots; \boldsymbol{x_N})^{\boldsymbol{T}}$, and the covariance matrix of the variable set is $\sum$. The determinant and the inverted matrix of the covariance matrix $\sum$ are $\det(\sum)$ and $\sum^{-1}$, respectively. $\sum$ will be a unit matrix when all the variables in $\boldsymbol{x} = (\boldsymbol{x_1}; \cdots; \boldsymbol{x_N})^{\boldsymbol{T}}$ are independent.

Table 1. Water environment items of teacher data

Category (Number of Items)	Water Environment Item	
Time of Sampling (4)	Year	Month
	Day	Hour
River/Flow Conditions (17)	Place of Sampling	Weather
	Water Level	Quantity of Flow
	Total Water Depth	Water Depth of Sampling
	Temperature	Water Temperature
	Vertical Visibility	Horizontal Visibility
	Water Smell	
	Time of Low Tide of Sampling Day	
	Time of High Tide of Sampling Day	
	Visual Appearance:	
	Water Color	Flow Strength
	Turbidity (Muddiness)	Floating Waste/Garbage
Watershed Conditions (7)	Length of Main Stream	Catchment Area
	Catchment Population	Number of Tributaries
	Annual Average Stream Flow	Number of Dams
	Number of Hydraulic Power Plants	
Water Quality Indexes For The Living Environment (10)	pH	BOD
	COD	SS
	DO	Saturation Degree of DO
	Total Coliform	
	The Amount of N-Hexane Extract (Oil)	
	Total Nitrogen	Total Phosphorus
Water Quality Indexes About Human Health (9)	Cadmium	Cyanogen
	Lead	Hexavalent Chromium
	Arsenic	Total Mercury
	Alkyl Mercury	PCB
	Dichloromethane	
Water Quality Index For Inflow Of Domestic Wastewater (1)	Ammonium Nitrogen	
Others (10)	Chromaticity	Turbidity
	Evaporation Residues	Total Hardness
	Potassium Permanganate Consumption	
	Sodium	Iron
	Manganese	Aluminum
	Residual Chlorine	
(7 categories in total)	(58 items in total)	

Table 2 Water environment quality standards for rivers [9]

Item Class	Water Use	Standard Value				
		Hydrogen-ion Concentration (pH)	Biochemical Oxygen Demand (BOD)	Suspended Solids (SS)	Dissolved Oxygen (DO)	Total Coliform
AA	Water supply class 1, conservation of natural environment and uses listed in A-E	6.5≤pH≤8.5	≤1 ㎎/L	≤25 ㎎/L	≥7.5 ㎎/L	≤50MPN/100mL
A	Water supply class 2, fishery class 1, bathing and uses listed in B-E	6.5≤pH≤8.5	≤2 ㎎/L	≤25 ㎎/L	≥7.5 ㎎/L	≤1000MPN/100mL
B	Water supply class 3, fishery class 2, and uses listed in C-E	6.5≤pH≤8.5	≤3 ㎎/L	≤25 ㎎/L	≥5.0 ㎎/L	≤5000MPN/100mL
C	Fishery class 3, industrial water class 1, and uses listed in D-E	6.5≤pH≤8.5	≤5 ㎎/L	≤50 ㎎/L	≥5.0 ㎎/L	-
D	Industrial water class 2, agriculture water, and uses listed in E	6.0≤pH≤8.5	≤8 ㎎/L	≤100 ㎎/L	≥2.0 ㎎/L	-
E	Industrial water class 3 and conservation of environment	6.0≤pH≤8.5	≤10 ㎎/L	Floating matter such as garbage should not been observed	≥2.0 ㎎/L	-

To apply the above multivariate distribution theory to water environment evaluation based on multiple water quality indicators, we define an evaluation score $ES(\mathbf{x}(0))$ as the probability by that a specific water quality level $\boldsymbol{x}(0) = (x_1(0); \cdots; x_N(0))^{\boldsymbol{T}}$ is realized. This definition is formulated as follows.

$$ES(\mathbf{x}(0)) = \int_{-\infty}^{x_1(0)} \cdots \int_{-\infty}^{x_N(0)} \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma) dx_1 \cdots dx_N \qquad (3)$$

The evaluation score (ES) defined in this way expresses how difficult it is to achieve a combination of current levels of each indicator in terms of their probability of realization. This realization probability is suitable for relative comparison of the difficulty of achieving the current level of the random variables, and is an appropriate indicator for ranking the current level of the random variables, i.e. the five evaluation goal variables in this study.

Table 3 shows the Evaluation Score (ES) and the ranking based on ES for the 104 Class-A rivers, as well as the ranking when each indicator of pH, BOD, SS, DO, and Total Coliform (TC) was evaluated separately.

Table 3 The evaluation score (ES) and the ranking results by ES and individual water quality indicators

ES	Ranking by						ES	Ranking by					
	ES	pH	BOD	SS	DO	TC		ES	pH	BOD	SS	DO	TC
0.995	1	1	2	2	1	1	0.589	53	59	51	49	49	55
0.978	2	5	1	1	4	3	0.573	54	53	55	55	55	51
0.976	3	3	4	7	3	4	0.570	55	46	54	51	56	58
0.976	4	4	5	4	5	6	0.567	56	49	52	56	53	56
0.966	5	10	3	3	2	2	0.550	57	63	57	57	57	53
0.964	6	7	8	8	6	9	0.547	58	50	58	60	60	61
0.957	7	13	6	5	8	8	0.546	59	54	62	61	62	60
0.954	8	2	9	10	9	10	0.515	60	64	59	65	58	62
0.952	9	16	7	9	7	11	0.507	61	60	60	59	61	63
0.943	10	11	11	12	11	7	0.491	62	61	61	58	59	59
0.940	11	8	12	11	12	5	0.487	63	65	63	63	63	57
0.938	12	20	10	6	10	12	0.466	64	57	65	67	66	71
0.928	13	6	15	14	13	17	0.458	65	62	68	64	68	67
0.920	14	12	14	18	14	19	0.448	66	67	64	66	65	64
0.916	15	14	13	13	15	13	0.441	67	66	66	62	64	72
0.912	16	9	20	16	18	26	0.435	68	73	67	68	67	65
0.901	17	15	17	17	19	14	0.431	69	68	71	70	70	66
0.891	18	17	16	19	16	20	0.430	70	70	69	71	69	75
0.886	19	24	18	21	17	21	0.408	71	77	73	74	73	69
0.876	20	18	19	15	20	24	0.395	72	75	70	72	72	68
0.872	21	26	21	22	21	15	0.394	73	74	72	69	71	76
0.864	22	23	23	26	22	28	0.375	74	69	78	77	78	78
0.846	23	27	22	20	23	22	0.369	75	71	74	80	74	73
0.837	24	19	24	24	24	30	0.364	76	85	75	73	76	70
0.834	25	21	25	27	26	23	0.363	77	76	76	75	75	85
0.833	26	30	27	28	25	16	0.353	78	72	77	76	77	82
0.829	27	22	30	25	30	25	0.352	79	81	79	79	81	79
0.827	28	31	28	32	27	18	0.340	80	78	80	78	82	86
0.824	29	28	29	30	28	31	0.336	81	87	82	83	80	74
0.796	30	33	26	23	29	34	0.312	82	79	83	84	88	90
0.783	31	29	31	31	33	27	0.299	83	80	86	82	86	80
0.765	32	25	33	33	32	40	0.292	84	88	85	87	83	84
0.756	33	39	32	29	31	33	0.291	85	91	81	81	79	87
0.756	34	36	36	35	35	38	0.271	86	86	84	86	85	83
0.756	35	37	37	41	36	29	0.253	87	82	88	85	87	91
0.753	36	32	34	40	34	42	0.247	88	94	87	91	84	77
0.730	37	41	35	38	39	37	0.229	89	83	89	89	89	92
0.725	38	35	38	36	37	41	0.222	90	92	90	90	90	81
0.698	39	47	39	34	40	32	0.213	91	84	92	88	93	96
0.698	40	42	42	42	42	36	0.163	92	93	93	96	91	93
0.696	41	40	40	45	38	35	0.150	93	96	91	93	92	95
0.683	42	34	41	37	41	45	0.124	94	89	97	98	95	94
0.681	43	51	43	39	43	39	0.113	95	101	94	97	94	88
0.677	44	38	45	46	44	48	0.088	96	102	95	92	97	89
0.675	45	45	46	44	45	46	0.080	97	90	99	95	96	97
0.675	46	55	44	43	46	44	0.068	98	100	98	94	100	99
0.664	47	43	47	47	47	43	0.050	99	99	96	99	98	100
0.651	48	48	48	50	48	52	0.044	100	103	100	102	99	101
0.633	49	56	50	48	52	47	0.038	101	98	102	101	102	98
0.627	50	52	53	52	50	50	0.037	102	97	101	100	101	104
0.622	51	58	49	54	51	54	0.013	103	95	103	104	103	102
0.617	52	44	56	53	54	49	0.005	104	104	104	103	104	103

Since different rankings are obtained by different indicators, we calculated two different kinds of Average Deviation (AD) between the two rankings $R_a(i)$ and $R_b(i)$ as follows to compare the ranking results.

$$AD_a(i) = \frac{\sum_b |R_a(i) - R_b(i)|}{5} \quad (4)$$

$$AD_{a,b} = \frac{\sum_{i=1}^{104} |R_a(i) - R_b(i)|}{104} \quad (5)$$

Where (i) indicates river number, and a, b are different water quality indicators. $AD_a(i)$ is the average deviation of river (i)'s ranking by indicator a from all the other five indicators, and $AD_{a,b}$ is the average deviation between two different water quality indicators for all 104 rivers.

$AD_{a,b}$ is shown in Table 4. The average value of the total deviation from all other rankings is shown in the last row of each column of the same table, and visualized as in Fig. 3. Clearly, the rankings based on pH and TC differ significantly from other ranking results and lack representativeness as a comprehensive evaluation. The assessments based on ES, BOD, and DO are relatively closer, with the evaluation based on ES being the most representative.

Table 4 Comparison of ranking deviation $AD_{a,b}$ between different ranking indicators

	ES	pH	BOD	SS	DO	TC
ES	0	3.90	1.29	2.13	1.31	3.46
pH	3.90	0	4.48	4.56	4.48	6.21
BOD	1.29	4.48	0	2.08	1.10	3.79
SS	2.13	4.56	2.08	0	2.38	4.10
DO	1.31	4.48	1.10	2.38	0	3.79
TC	3.46	6.21	3.79	4.10	3.79	0
Average AD	2.42	4.73	2.55	3.05	2.61	4.27

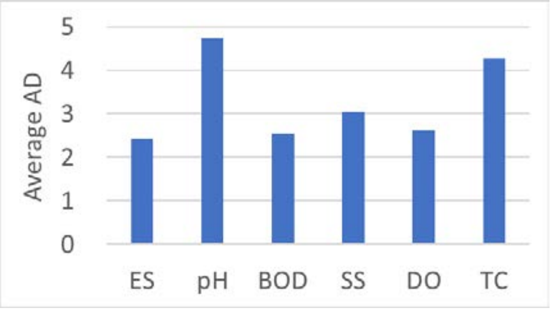

Fig. 3 Average $AD_{a,b}$ of different ranking indicators for all rivers

The following can be derived from the ranking results above.

First, the ranking results by different indicators and methods are different but quite close (Table 3). There was no significant reversal in the rankings by different indicators.

Next, as shown in Table 4, the evaluation based on ES has the smallest average deviation of 2.42 from the other ranking results, and it can be said that the ranking based on ES is relatively the most reliable of the compared models.

Furthermore, among the rankings based on single indicators, the evaluation based on BOD is the best, followed by the evaluation based on DO. On the other hand, the rankings based on pH and TC have larger deviations from the other evaluation results than BOD and DO, which are due to the characteristics of these two indicators themselves that both pH and TC are much easier to be affected by a single rainfall or urban drainage system overflow event.

4.3 Training of Neural Network Model

The neural network model has been trained (put under a learning process) with the teacher data that composed with the evaluation score (ES) as the only evaluation goal variable, and all the other 58 environment items as the explanation variables. The training process is based on the learning procedures but it is still a process of trial and error because there are still many details that remain undecided, such as a suitable step size of optimization, a suitable output function, an efficient order to present the teacher data to the neural network model, and a proper initial network size (layers and units in each layers). The learning process was stopped after the trained neural network model is able to reproduce the entire teacher data with an acceptable error, which was set in this study to be below 2% in mean relative error for the evaluation goal variable.

5. IDENTIFYING IMPROVEMENT MEASURES

5.1 Sensibility Analyses of Explanation Variables

We need to define what is the most effective measures of river environment improvement to identify the effective measures. In this study, the effectiveness of an environment improvement measure is defined with the sensibility of an explanation variable to the evaluation score (ES).

The well-trained neural network model has been used to carry out a sensibility analysis for all the explanation variables to identify how much each variable impacts on the evaluation scores. The sensibility coefficient of an explanation variable is defined as the partial derivative of the evaluation score regarding each explanation variable as follows.

$$S_i = \frac{\partial ES(X_1, X_2, \cdots, X_N)}{\partial X_i} \Big|_{at\ a\ given\ variable\ value\ set} \quad (6)$$

where S_i is the sensitivity coefficient of variable X_i at a given variable value set $\boldsymbol{x}(0)$

5.2 Identifying Class-A Rivers as a Whole

For all the variables, the sensitivity coefficients have been evaluated by treating the 104 major rivers as a whole set for the average variable value of the last data year, 2018. These sensitivity analyses have identified seven variables that have the most significant and meaningful impacts on the evaluation score (ES), as shown in Table 5. As for all the other variables, the sensitivities were not great enough to treat them as variables with a considerable impact in terms of average variable values at the time of the year 2018. We consider these results to be reasonable and reliable, as they are largely consistent with previous research results using different evaluation goal variables [12].

The environment items with the greatest impacts on the achievement degree of water environment standards can be classified into three categories: natural factors, human factors and mixed factors.

Natural factors include month of sampling and annual average stream flow. The fact that the month of sampling has a great impact means that the water environment has a clear tendency to seasonal change. This is because of the subtropical climate pattern in Japan with a clear rainy season and a typhoon season. Seasonal rainfall change is considered the main cause for the seasonal change tendency in a river water environment. This is consistent with the fact that the annual average stream flow is ranked as the second most important factor in the water environment. This is also a reasonable result expected from the common sense that water quantity has a tremendous impact on water quality.

Human factors include catchment population, total nitrogen, dichloromethane, and the number of hydraulic power plants. Catchment population implies that human activity is one of the main factors that can make a significant difference in water environments. Total nitrogen in the water environment is mainly contributed by sewage inflow and agriculture drainages (overuse of fertilizers), and dichloro-methane is almost entirely from industrial wastewater. Power plants are the most controversial factor. The conflicts between power generation and the conservation of the river environment have been problems in most basins, and a priority problem of river flow for different purposes remains to be resolved.

The only mixed factor is turbidity or muddiness in terms of visual appearance. Visual turbidity usually gets remarkably worse right after a heavy storm that causes landslides or debris flow in mountain areas, soil erosion in farmland and sewage overflow in urban areas. Both catchment natural condition and farming or urban human activities are contributing indirectly to river flow turbidity during rainy time.

5.3 Identifying Individual Rivers

Efforts were also made to identify the most effective water environment improvement measures for individual rivers. Table 6 presents the top five environmental improvement measures that were identified for rivers A and B. It should be noted that River A is a major river with a core city in its middle reaches and a metropolis downstream, and its current

Table 5 The environmental items with the greatest impacts on Evaluation Score (ES)

Environment Item	Sensitivity	Descriptions
Month of Sampling	0.28	It means that the water environment has a clear tendency to seasonal change.
Annual Average Stream Flow	0.16	It is reasonable and expected that water quantity has a tremendous impact on water quality.
Catchment Population	0.13	This just reconfirmed that human activity is one of the main factors that are able to make a great difference on water environments.
Visual Appearance: Turbidity (Muddiness)	0.09	Visual turbidity usually gets remarkably worse right after storms that cause landslides or debris flow in mountain areas, soil erosion in farmland, and sewage overflow in urban areas.
Total Nitrogen	0.08	Total nitrogen in the water environment is mainly contributed by sewage inflow and agriculture drainages.
Dichloromethane	0.03	Dichloromethane is almost entirely from industrial wastewater.
Number of Hydraulic Power Plants	0.02	Power plants can change river flow and, consequently water environments

Table 6 The most effective environment improvement items for individual rivers

Ranking of effectiveness	River A	River B
1	Month of Sampling	Month of Sampling
2	BOD	Visual Appearance: Turbidity
3	Visual Appearance: Turbidity	Total Nitrogen
4	Catchment Population	BOD
5	DO	Vertical Visibility

water quality ranking is among the worst 40 of the 104 Class-A rivers. River B, on the other hand, is considerably polluted with industrial wastewater and livestock industry drainage, and its current water quality ranking is among the worst 10.

In both river A and river B, the season continues to be the strongest factor causing changes in water quality. Subsequently, the most significant influencing factor is the Visual Appearance of the water quality (turbidity), and improvements in this aspect are expected to greatly contribute to the enhancement of the river environment. In river A, the population of the watershed, along with BOD (Biochemical Oxygen Demand) and DO (Dissolved Oxygen), are significantly involved. This implies that improving the domestic wastewater of urban residents in the middle and lower reaches, such as through advanced treatment, would have a substantial effect on river environment improvement. For river B, the reduction of total nitrogen has a significant effect on water quality improvement, indicating that the drainage from the livestock industry and farming within the watershed is the cause of the current deterioration in water quality.

6. CONCLUSION

In this study, an artificial intelligence model was constructed to discover the most effective river environmental improvement measures. This AI model outputs an evaluation score (ES) that determines the river environmental ranking among a given river group, which integrates five water quality indicators that are used in the river environmental standards and utilizes 58 river environmental items as explanatory variables. Using this model together with a sensitivity analysis of each environment item to the evaluation score, the most critical environmental items for improvement were identified for 104 Class-A Rivers in Japan, and the most effective river environmental improvement measures were also discovered for individual rivers. These improvement measures are in complete alignment with the current state of the rivers and are highly rational.

This model will be used to identify effective river environmental improvement measures for each of the remaining rivers, and to explore their applicability and challenges by comparing with the traditional improment measure selection methods. The results of this study are expected to significantly contribute to the future development of river environmental planning methodologies.

7. REFERENCES

[1] Minister of Land, Infrastructure, Transport and Tourism, Annual report on the water environment details of the Class-A Rivers of Japan, 2024, pp.1-356.

[2] Hagihara Y. and Hagihara K. ed., Planning of Urban Water and Green Space, Kyoto University Press, 2010, pp.225-232.

[3] Urban Science Study Group, ed., Introduction of Urban Science, Soseisya Publisher, 2020, pp.35-50.

[4] Zhang S. P. and Kido Y., A Study on the Environment Evaluation Method for Aise River and the Effectiveness of River Environment Improvement Measures, Urban Science Studies, No. 21, 2016, pp. 45-56.

[5] Zhang S.P. and Qi Jie, Evaluating Habitability of Water Environments for Fireflies with an Artificial Intelligence Model, International Journal of GEOMATE, April, 2020, Vol.18, Issue 68, pp. 87-93.

[6] Asou H., The Information Processing by Neural Network Models, Sangyo Publisher, 1988, pp.1-286

[7] Rumelhart D. E., Hinton G. E., and Williams R. J., Learning Representations by Back-propagating Errors, Nature, Vol. 323, No. 9, 1986, pp. 533-536.

[8] Ministry of Land, Infrastructure, Transport and Tourism of Japan, Water Information System, http://www1.river.go.jp/ (an Online Open Source Database).

[9] Ministry of Environment of Japan, Environmental White Paper, 2020, pp.1-426.

[10] Thomas A. Severini, Elements of Distribution Theory, Publisher: Cambridge University Press, 2005, pp.1-515.

[11] Stein, C., A Bound for the Error in the Normal Approximation to the Distribution of a Sum of Dependent Random Variables, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. 1972, Vol.6, No.2, pp.583-602.

[12] Zhang S.P. and Qi Jie, A Sensitivity Analysis of River Environment Factors Through Deep Learning, International Journal of GEOMATE, Sept. 2022, Vol.23, Issue 97, pp.146-153.