Article

Applicability Evaluation of the Hydrological Image and Convolution Neural Network for Prediction of the Biochemical Oxygen Demand and Total Phosphorus Loads in Agricultural Areas

Chul Min Song ^{1,*} and Jin Soo Kim ²

¹ Department of Policy for Watershed Management, The Policy Council for Paldang Watershed, Yangpyeong 12585, Korea

² Department of Agricultural Rural Engineering, Chungbuk National University, Choengju 28644, Korea; jskim@cbnu.ac.kr

* Correspondence: kaih4@nate.com; Tel.: +82-01-3117-3817

Received: 18 September 2020; Accepted: 3 November 2020; Published: 5 November 2020

Abstract: This study employed a convolution neural network (CNN) model, hitherto used only for solving classification problems, with two-dimensional input data to predict the pollution loads and evaluate the CNN model's applicability. A CNN model generally requires two-dimension input data, such as photographs in previous studies. However, this study's CNN model necessitates the numerical images that reflect hydrological phenomena due to the nature of the study. A hydrological image was used as the input data for the CNN model in this study to address this issue. The last layer of the CNN model was also transformed into a linear function to derive the continuous variable. As a result, the Pearson correlation coefficient, which represents the relationship between the measured and predicted values, demonstrated a Biochemical Oxygen Demand (BOD) load model of 0.94 and a Total Phosphorus (TP) load model of 0.87. Nash–Sutcliffe efficiency was used to evaluate the model performance; the BOD load model was 0.83, while the TP load model was 0.79, respectively, indicating good performance. These results demonstrate that the hydrological images led to stable model learning and generalization, and the proposed CNN model is suitable for predicting the pollution load, with potential future applications in various fields.

Keywords: convolution neural network; hydrological image; biochemical oxygen demand load; total phosphorus load; prediction

1. Introduction

Until the end of 1990s, South Korean river water quality management policy was managed by setting a certain level of standard concentration, which was observed in most major rivers. However, in the 2000s, the level of pollution in rivers gradually got worse, and the government changed the paradigm of river management from qualitative to quantitative management, and set the pollution load as the river management standard.

The most important part in establishing a quantitative management plan is the prediction of pollution load in the management watershed, which includes two methodologies, a conceptual and empirical model. The conceptual distribution model is a model that mathematically expresses spatial characteristics, such as variations in human activities or natural conditions, within a watershed. However, the conceptual distribution models mainly used in previous studies, such as Soil Water Assessment Tool [1], Areal Non-point Source Watershed Environment Response Simulation [2], Agricultural Non-Point Source Pollution Model [3], and Storm Water Management Model [4],

have disadvantages in that they require vast data including human activity, meteorological and spatial data, and the model cannot be operated without meeting such requirements. The reliability of a conceptual model can be evaluated by two different methods: sensitivity analysis to understand the response of the result using the variation in the input value, and uncertain analysis to determine the influence degree that the uncertainty of the input value propagates to the result value. Although the conceptual model has the advantage of being able to explain a clear causal relationship, it requires a vast amount of input data and cumbersome post-processing processes, such as the inspection and correction works, to adjust the gap between the measured and simulated data, as well as the reliability evaluation of the model. The methodology of the empirical model could be helpful in a decision-making process that leads to reasonable conclusions from collected data. The model using this methodology is characterized as a model developed using a stable relationship between the input and result data without using physical laws, as an alternative to the conceptual model [5]. Choosing an appropriate statistical method with the ability to appropriately use and interpret data enables accurate analysis of research results, leading to an efficient and valuable research by reducing unnecessary trial, error, and data loss in the overall research: from planning to subject selection, conduct of the research, data input and analysis, as well as recording and interpretation of the results. Additionally, the empirical model, unlike the conceptual one, is considered to be more efficient as the data range can be determined arbitrarily according to the subjectivity of the researcher, and is relatively simple to develop and operate. Especially in establishing a watershed management plan, using a simple model is more efficient than dealing with a complex model in a field with lack of experts.

In this regard, an artificial neural network (ANN), which belongs to the category of empirical models, used to be neglected by the academic world in the past. However, with the recent advances in big data and computer hardware, results demonstrating a dramatic improvement in ANN prediction and performance are being reported, continuously attracting active researchers in this field. Moreover, studies on existing ANN structures are also transferring into deep learning with complex layers of neural networks or a deep neural network (DNN), obtaining a dramatic increase in accuracy. While researchers that prefer the conceptual model suggest that using an empirical model such as ANN is not suitable for model development since the dynamic relationship between the input and result data cannot be clearly explained [6], many previous studies have reported excellent results through an ANN function that can grasp the nonlinear relationship of data to predict the pollution load [7–11]. Some researchers have reported that it is even possible to detect physical processing in a hydrological model using an ANN, with excellent results [12–14].

Most previous studies, such as the aforementioned pollution load prediction model, remain at the level of optimizing the relationship between the input and result data consisting of one-dimensional vectors, or the parameter estimation of a regression model based on ANN. However, considering the current development trend of ANN, most of these ANN approaches on pollution load prediction are unable to keep up with the technology level of the current ANN; fail to incorporate various ANNs; or focus on improving the performance through optimizing the ANN structure by adjusting the number of hidden layers and nodes of the hidden layer in the ANN, or by employing a learning method for ANN according to the type of target problem or a specific learning method suggested by a researcher. However, the current development trend of ANN or DNN, such as the aforementioned prior studies, is mainly conducted by adjusting the number of hidden layers or nodes of hidden layers within ANN to optimize the ANN structure or by employing a learning method for ANN according to the specific learning method suggested by a researcher. These have limitations that do not reflect the application or response of DNN, where radical technology development takes place.

One of the deep learning algorithms, the convolution neural network (CNN), can be proposed to improve this. CNN has the performance of extracting spatial and geometric characteristics through the input data of two- or three-dimensional structures, as well as one-dimensional structures, so it is used in various fields such as image classify analysis and signal processing. In this regard, the one-direction CNN model (1DCNN), which uses input data of one-dimensional structures, has been reported in the

application of research of a rainfall-runoff model [15,16], but these are also at the level of the existing ANN model.

Therefore, this study aims to develop a pollution load prediction model using the extract function of the CNN model and evaluate the performance of this study's model. The CNN model requires images as input data for model learning; to this end, this study attempts to use hydrological images as the input data and describe its applicability for the CNN model's input data.

2. Materials

2.1. Study Area

The study area is Yangpyeong-gun, Gyeonggi-do, South Korea, and is referred to as Heukcheon (E 127°31'55.39", N 37°27'37.69") in the Han River watershed among the unit watershed classified for the Total Maximum Daily Load System (TMDL) (Figure 1). The TMDL is a national policy of watershed management system currently enforced in South Korea to achieve the purpose of efficient land use of the national territory and water quality conservation. This system has establishment and management plans for the discharge and reduction of pollutant loads and land use for the unit watershed. Heukcheon is a watershed where the river called Heuk-River; it originates from Seongjibong, Cheongun-myeon, Yangpyeong-gun, with no rivers flowing in from other administrative districts or watersheds. The target area is a watershed adjacent to Seoul and incorporated into the metropolitan area, with Heuk-River flowing into Lake Paldang, formed by the Paldang Dam. Since Lake Paldang is a water resource used by approximately 50% of the population of South Korea, it is a critical resource for national security. Due to these features, the central government of South Korea has imposed strict regulations and has declared it as water resource protection area, waterside area, special countermeasure area, and nature conservation zone, to maintain the health and stability of the water quality of Lake Paldang.

Figure 1. Schematic of study site.

As such, the study area has been maintained as a rural area with little variation in land use. The area is 314.0 km², the river length is 42.9 km, and the average slope is 27°, which consists of 79% forests and 15.2% agricultural land (Table 1).

Table 1. Summary of study area.

Landcover	Water	Urban	Barren	Pasture	Forest	Paddy	Upland	Total
Area (km ²)	8.3	10.9	7.1	1.4	238.3	26.0	22.0	314.0
Proportion (%)	2.7	3.5	2.3	0.4	75.9	8.3	7.0	100.0

2.2. Data Collection

In this study, the amount of data used was kept to a minimum to build a simple model. Only four types of data, precipitation, discharge, land use map, and soil map, were collected from 2006 to 2019; water quality data were used for this study as well. All the necessary data of this study are available online: the precipitation data from the Meteorological Administration [17]; the discharge data and the soil map from the Water Resources Management Information System [18]; the land use map from the Environmental Geographic Information Service [19]; and the water quality data from the National Institute of Environmental Research [20].

The discharge data are daily data measured at Heukcheon Bridge, as shown in Figure 1, while the precipitation data are daily data measured at six precipitation gauges located inside and outside the study area, as can be observed in Figure 1. The land use and soil maps are in grid format. The resolution of the data is 30 × 30 m, with a scale of 1:5000. Land uses are classified into seven different categories: water, urban, paddy, meadow, forest, upland, and barren. The soil map is a data with 59 physical properties. Two types of water quality data, Biochemical Oxygen Demand (BOD) and Total Phosphorus (TP), measured every eight days at the same point as the discharge measuring point, were collected and used in the model. The reason for selecting BOD and TP was that they are management items in TMDL implemented in South Korea. In regions under strict regulations such as Heukcheon A, the increase or decrease in BOD and TP loads are closely associated with regional development, which makes it a highly sensitive factor for the local community. Therefore, the prediction of BOD and TP loads is more important than in any other areas.

2.3. Method of Study

The CNN model employed to predict the pollution load in this study is mainly used for the classification problem, with supervised learning as the learning method. Supervised learning requires both input and target data, while the CNN model uses images as the input data and their corresponding labels for learning. In general, the CNN model has the characteristic of learning the relationship between target values by extracting the color value or geometric characteristic of the image. However, this cannot reflect the characteristics of the factors that affect the pollution load in our case, which invalidates the premise of the empirical model. Even if training between the general image and target value proceeds smoothly and yields excellent results, they become meaningless as the relationship between the input data and the result value cannot be established. Therefore, here, we set the meteorological, topographical, and hydrological conditions as the factors influencing the pollution load while constructing the input data as recognizable images in the CNN model. Subsequently, it was used to train the CNN model for the nonlinear relationship between image and pollution load, while the suitability and performance of the model were evaluated through the pollution load prediction.

2.3.1. Hydrological Image

The CNN model uses an algorithm that solves a problem by extracting characteristics, such as color values and locations, in a two-dimensional (N × M) space by receiving RGB (red, green, blue) or gray scale images as input data. This study also requires images as the recognizable input data for

CNN model, but due to the nature of the study, numerical images reflecting hydrological influences must be used rather than general photographic data. For that, a phenomenon in which BOD and TP runoff differ from each other due to the influence of land use and soil characteristics when rainfall occurs, was adopted to create an image. To implement this efficiently, this study used hydrological images, which is a concept proposed by Song [21], and is defined as a set of grids with dimensionless hydrological properties in a two-dimensional space for an arbitrary watershed.

Hydrological images are produced using the hydrological curve number (CN) published by the Soil Conservation Service (SCS), the predecessor of the National Resource Conservation Service (NRCS). CN is a method that can evaluate hydrological effects through direct runoff estimation based on conditions such as precipitation, land use and soil map [22,23]. Since pollutants such as BOD and TP are directly affected by the discharge rate, the hydrological images created using CN, which can simulate direct runoff, are regarded as suitable for the pollution load simulation in this study.

Generally, as a CN-based prediction method, the direct runoff is estimated by calculating CN that represents the entire watershed, or by allocating CN using an area-weighted mean. However, Song [21] reported that this method only generates very monotonous images, which cannot be used as input data for CNNs that require various learning materials, and suggested assigning CN to each grid of 30×30 m, similar to pixel, the basic unit of a picture, to create a hydrological image. Accordingly, this study attempted to assign CN to each grid of 30×30 m, and generate an image in which each grid variation is according to precipitation conditions to obtain an effect similar to that of a photograph.

To allocate CN for each grid, the land use and soil maps must possess the same resolution, while precipitation data must have the same resolution in a two-dimensional structure to incorporate precipitation conditions into each grid. While the land use and soil map data are of the same resolution, precipitation is recorded as one-dimensional data at each precipitation gauge and requires conversion into two-dimensional data. Traditionally, the Thiessen polygon method is proposed as a method of representing precipitation data in two dimensions. However, it suffers from a disadvantage of causing distortion or discrepancy from actual precipitation due to discontinuity of numerical variation at polygonal boundaries. Therefore, here, precipitation was expressed as two-dimensional data using inverse distance weighting (IDW) instead of the Thiessen polygon method. Since IDW is an interpolation method based on Tobler's law, it enables estimation of continuous precipitation variations, thereby compensating for the discontinuity caused by the Thiessen polygons. In this study, as shown in Figure 1, precipitation data from six precipitation gauges were used as input data for the IDW tool of ArcGIS 10.0 to generate precipitation data in a grid format with a resolution of 30×30 m.

Equations (1) and (2), suggested by SCS, were used as the method of calculating the runoff of each grid. However, the SCS proposes to consider the initial loss (I_a) in Equations (1) and (2). I_a is defined as the quantity that is not released as direct runoff by the interception of plant-borne on the ground, infiltration (or percolation) into underground, or evaporation during the event. The SCS explains that, theoretically, all precipitation minus I_a represents the amount of discharge in a particular watershed. In addition, the SCS suggested that I_a set to $0.2 S$

$$Q = \frac{(P - 0.2S)^2}{(P + 0.8S)} (P > I_a, \dots I_a = 0.2S) \quad (1)$$

$$Q = 0 (P < I_a, I_a = 0.2S) \quad (2)$$

$$S = \frac{25400}{CN} - 254 \quad (3)$$

where Q means direct runoff (mm), but here, it denotes the attribute value of the hydrological image, P denotes precipitation (mm), S denotes the potential watershed storage (dimensionless) and I_a presents the initial abstraction (mm). Although CN value is provided by SCS, it is not suitable for use in South Korea, with different topographic conditions. We modified CN to fit the topographic conditions of South Korea based on the Design Flood Estimation Techniques [24].

CN in Equation (3) corresponds to the AMC_{II} condition, and the discharge calculation without considering the AMC condition results in a large error between the measured and predicted values.

Here, AMC means antecedent soil moisture condition (AMC). AMC has three levels, such as AMC_I , AMC_{II} , and AMC_{III} . AMC is to reflect the phenomenon of surface runoff being different depending on the soil's moisture condition by precipitation. In this regard, SCS proposed the concept of the 5-day antecedent cumulative precipitation (P_5) to judge a level of AMC and suggested adjusting CN according to variation in AMC by P_5 following dry and wet seasons. CN_I and CN_{III} were adjusted by Equations (4) and (5), respectively. In this study, considering the characteristics of South Korean climate, the wet season was set from June to September and the dry season from October to May, such as in Table 2

$$CN_I = \frac{4.2CN_{II}}{10 - 0.058CN_{II}} \sum_{i=1}^n P_i < 12.7 \text{ (or 35.6)mm} \quad (4)$$

$$CN_{III} = \frac{23CN_{II}}{10 + 0.13CN_{II}} \sum_{i=1}^n P_i > 27.9 \text{ (or 53.3)mm} \quad (5)$$

where CN_I refers to the value at AMC_I and CN_{III} refers to the value at AMC_{III} . As a final step, for the aforementioned reasons, the image of the non-precipitation condition was discarded as it was just a fixed CN value, by collecting only the precipitation data. The collected data were converted into a TIF file, one of the image format files, while maintaining the same two-dimensional structure as the land use and soil maps, and constructed as CNN input data. The reason for using the TIF format is that other formats, such as PNG, BMP, and JPG, may result in data loss as the runoff information, which is an attribute value of the grid, is converted into a new different value during the conversion to a color value in the range of 0 to 255 integers.

Table 2. Classification of antecedent soil moisture condition.

Antecedent Soil Moisture Condition	P_5 (mm)	
	Dry Season	Wet Season
AMC_I (Dry condition)	$P_5 < 12.7$	$P_5 < 35.6$
AMC_{II} (Normal condition)	$12.7 \leq P_5 \leq 27.9$	$35.6 \leq P_5 \leq 53.3$
AMC_{III} (Wet condition)	$P_5 > 27.9$	$P_5 > 53.3$

The process of constructing a hydrological image is illustrated in Figure 2, such as estimating the 5-day preceding cumulative precipitation for each grid, reflecting the dry and wet season conditions, adjusting CN by AMC, and calculating outflow by grid. The work environment used was Python 3.7 [25], an open source language.

Figure 2. Schematic of hydrological image construction process.

2.3.2. Target Data

Target data are used to train the model as the desired output for the given input data given in the neural network model. This means the supervised learning among the method of model training, which aims to decrease the error between the predicted and target data while continuously improving the weight within the neural network. Generally, the target data were applied in model training and testing, here the testing means to predict pollutant loads and to evaluate the model. The target data consist of BOD and TP loads for the whole study area in this paper. These pollution load data corresponding to the date of the generated hydrological image were used as the target data and were constructed as a CSV format file. Here, the pollution load was calculated by water quality and discharge for the corresponding date of the hydrological image.

2.3.3. CNN Structure Configuration

The CNN, a model that mimics objects-recognizing organisms' visual processing process, is facing the opportunity for further development by LeCun et al. [26]. LeCun et al. [26] proposed the LeNet-5 structure, as shown in Figure 3, and reported very high performance in recognizing the handwriting on the check. The LeNet-5 structure consists of the repetition of a convolution layer and a pooling layer, adding a fully connected layer. In recent years, several researchers have transformed LeNet-5's structure to compete for the model's performance in competitions such as ImageNet Large-Scale Visual Registration Challenge (ILSVRC).

Figure 3. Original image published of LeNet-5 architecture in LeCun et al. [26].

Unlike the existing ANN, which uses traditional one-dimensional input data, the CNN is a model that can process two- or three-dimensional input data, especially by maintaining the input data's spatial information, so that it can be able to grasp the spatial characteristics of the space.

CNN can be divided into two parts: Part 1 is a part that extracts image characteristics, a core function of CNN, consisting of the repetition of the convolution layer and the pooling layer.

Part 1 of CNN maintains the structure of the I/O data of each layer and recognizes the characteristics with adjacent images while maintaining the image's spatial information. Part 2 of CNN means a deep neural network consisting of several hidden layers or dense layers between the flatten and output layers and has a role in deriving the predicted results [27,28].

We attempted to revise the CNN structure to simulate continuous variables while maintaining its image extraction function, such as in Figure 4. Part 1 was designed in the most general structure in which the convolution layer and the max-pooling layer are alternately repeated five times. In general, Part 2 of CNN is called a fully connected layer, whose input is the image characteristics extracted in Part 1; it is connected to a hidden layer through a flattened layer that converts it into a one-dimensional structure, and finally derives the result value. To fulfil its original purpose of solving the classification problem, CNN adopts a softmax or sigmoid function as the activation function, and a categorical cross-entropy or binary cross-entropy function as its loss function. However, since the runoff value to be simulated is an unspecified continuous variable, the existing Part 2 structure is not appropriate for this study.

Figure 4. Structure and modified part 2 of the convolutional neural network (CNN) model for continuous variable prediction.

Therefore, here, three dense layers of Part 2's structure were arranged after the flattened layer; the activation function of the third dense layer was transformed to a linear function to derive the result of the continuous variable (Figure 4).

Additionally, it was designed in a structure in which a batch normalization layer was added to avoid overfitting after the second dense layer. Keras [29] with the backend of Tensorflow [30], a machine learning library released by Google, was used for all environments for CNN model design, implementation, and operation, with the implementation and experiments conducted in Python language. Keras is a Python-based high-level deep neural network application programming interface

(DNN API) which supports all special functions of Tensorflow and demonstrates high flexibility in implementing DNN models. It facilitates fast experiments with the efficient use of CPU and GPU [29].

2.3.4. Detailed Model Configuration

For convolution layer operation, a separate 2D plane function called kernel is used, which is a kind of image filter. In Keras, the kernel size can be set arbitrarily, and it was set to 3×3 for all convolution layers in our model. Similarly, the number of kernels of the five convolution layers were defined as 32, 64, 128, 256 and 512 in the order of layer. The activation function of all convolution layers was set as the rectified linear unit (ReLU) function (Equation (6)), which is commonly used in CNN models, and has been reported to have optimization efficiency without gradient vanishing [31,32].

$$f(x) = \max(0, x) \quad (6)$$

The pooling layer is a sub-sampling technique in which only important information is left while reducing the image size. Based on this, the computer memory can be used efficiently, while avoiding overfitting with reduced data volume for calculation. There are max-pooling and average pooling methods. Since average pooling may lead to data loss, max-pooling was used in this model. The size of the pooling can also be set arbitrarily. In this study, the convolution kernel size was set to 3×3 .

The CNN structure in this study is shown in Table 3, in which the number of total parameters (weights of the neural network) was 13,985,769, the number of trainable parameters was 13,985,513, and the number of non-trainable parameters was 256. The non-trainable parameters were caused by the batch normalization layer with 128 nodes. Finally, the activation function of the last dense layer is set as a linear function to derive continuous variables, with the number of nodes set to one to match with the target data.

Table 3. Summary of Study Model.

Convolution Layer	Output Shape (Raw_Size, Column_Size, Image_Channel)	Parameter (Weighted Value)	Activation Function
Conv2D_1	317, 478, 32	320	ReLU
Max-Pooling2_1	158, 238, 32	0	
Conv2D_2	79, 119, 64	18,496	ReLU
Max-Pooling2_2	39, 59, 64	0	
Conv2D_3	39, 59, 128	73,856	ReLU
Max-Pooling2_3	19, 29, 128	0	
Conv2D_4	19, 29, 256	295,168	ReLU
Max-Pooling2_4	9, 14, 256	0	
Conv2D_5	9, 14, 512	1,180,160	ReLU
Max-Pooling2_5	4, 6, 512	0	
Fully Connected Layer	Output Shape (number of node)	Parameter (Weighted value)	Activation function
Flatten_1	12,288	0	
Dense_1	1000	12,289,000	ReLU
Dense_2	1000	128,128	ReLU
Batch Normalization	128	512	
Dense_3	1	129	Liner
Total parameters: 13,985,769			
Trainable parameters: 13,985,513			
Non-trainable parameters: 256			
Optimizer function: Adam (learning ratio = 1×10^{-4})			
Loss function: MSE			See Equation (7).
Metrics: MAE			See Equation (8).
Epoch: 1000 iterations			

The neural network is optimized by updating the weights of the neural network in order to minimize the error between predicted and target values of models derived from the model training process. As previously mentioned in Section 2.3.2, this process means supervised learning among the model training methods, and the model of this study, the CNN model, applied the same learning method. Here, the error is called Loss in the field of neural networks, and the purpose of supervised learning can be explained as minimizing Loss. Changes in Loss can be used as indicators to evaluate the results of model learning, and changes in Loss can be monitored using the Loss function. The CNN model mainly has used Loss functions such as binary crossentropy, categorical crossentropy, and sparse categorical crossentropy to evaluate the result of model learning for classification purposes, which are not suitable for predicting continuous variables, such as this study. Hence, this study employed MSE (Equation (7)), which is widely used in regression issues as a function of Loss to evaluate the degree of model learning. In addition, the item to be considered with the model learning is the predictive performance of the trained model. As a metric to examine this, the study applied a mean absolute error (MAE), as shown in Equation (8)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2 \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y'_i - y_i| \quad (8)$$

where, y_i and y'_i denote the measured and predicted values, respectively. The closer the MSE and MAE to 0, the higher the model training.

Even if the results of the Loss function and the predictive performance of the model are good or excellent, it is only the result of training between the input data and the target data for model learning; therefore, the model must be verified regarding whether it draws a smooth prediction even for new data (unseen data) that are not used in model training. This is also called the generalization of the model, in this study, some of the collected data were selected as validation datasets to evaluate the generalization of the model. This study judged the result of the model learning using Loss, MAE, Validation Loss, and Validation MAE. Here, Loss and MAE mean metrics for input data, and Validation Loss and Validation MAE are for validation datasets.

In the field of neural networks, stochastic gradient descent (SGD) [33] is considered as the most basic methodology in terms of optimizer algorithm. It updates weights by obtaining a slope through differentiation and then updating towards a lower slope to reduce loss. Here, “stochastic” means the weights are not calculated at once, but rather some samples are obtained and calculated gradually. With a sufficient amount of learning, SGD works well, but has the disadvantage of being very noisy and slow. Recently, to improve its shortcomings, optimizer algorithms, such as Momentum [34], Nesterov Accelerated Gradient (NAG) [35], Adagrad [36], Nadam [37], AdaDelta [38], RMSProp [39], and Adam [40] have been developed to increase the learning speed and accuracy based on SGD. Adam, which stands for adaptive moment estimation, combines the concepts of momentum optimization and RMSProp. Like momentum optimization, it follows the exponential decaying average of the gradient, and like RMSProp, it adheres to the exponentially decremented mean of the square of the last gradient [41]. Owing to these characteristics, Adam enables stable and quick optimization of the model, and performs better than other optimization methods. Therefore, in this study, the weight optimization algorithm employs Adam; the learning rate was 0.0001 and the epoch was set 1000 times.

2.3.5. Evaluation of Model

To evaluate the model, its performance was determined by comparing the measured and predicted values, based on three criteria: Pearson correlation coefficient (r), Nash–Sutcliffe efficiency (NSE) and mean absolute percentage error (MAPE).

r is expressed in the range of -1 to 1 using Equation (9), and represents the linearity between the measured and predicted values. The closer r is to 1 , the stronger the linearity of the predicted and measured values of the model. As it converges to 1 , the model performance becomes better.

$$\gamma' = \frac{\sum_{i=1}^n (y_i - \bar{y})(y'_i - \bar{y}'_i)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 (y'_i - \bar{y}'_i)^2}} \quad (9)$$

NSE is presented as in Equation (10). If the result value is 1 , it means that the model result matches perfectly with the measured value. As it converges to 0 , the model performance degrades.

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

MAPE, as a percentage, is denoted as in Equation (11). The key of MAPE is not to perceive errors as absolute size, but to comprehend them as proportionate. With a range of 1 – 100 , the MAPE has the characteristic that the interpretation of the results is useful. The closer the MAPE to zero, the better the performance of the model

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y_i} \right| \quad (11)$$

where, y_i and y'_i denote the measured and predicted values, respectively, while $\bar{y}$ and $\bar{y}'_i$ indicate the average measured and predicted values.

3. Result and Discussion

3.1. Data Collection during the Period of Study

As mentioned in Section 2.3.1, the attribute value of each grid is assigned in the hydrological image by Equation (1) once rainfall occurs. During the period of study, there were 680 rainfall events. In this regard, the variation of discharge according to precipitation on the corresponding days of 680 rainfalls were shown in Figure 5.

Figure 5. Variation in precipitation and discharge during the period of study: (a) changes in precipitation and discharge; (b) scatter plot between precipitation and discharge.

As shown in Figure 5a, 680 rainfalls ranged from 2.2 to 271.6 mm, with the average daily discharge being 0.18 to 704.27 m³/s. The relationship between precipitation and discharge is shown in Figure 5b, and the two-sample assuming unequal variances test (t -test) showed $p < 0.01$.

In general, if the p -value of the t -test results in less than 0.05, the two groups are determined to be statistically significant. The p -value of Figure 5b shows under 0.01, which indicates a significant level of causality between precipitation and discharge data. In other words, precipitation and discharge data have a proportional relationship, statistically, as shown in the scatter and trend lines (red dotted lines) of Figure 5b.

The variation in pollution concentration was shown in Figure 6. The variation in BOD concentration according to the discharge was found to be from 0.2 to 6.4 mg/L (Figure 6a), and the variation of TP concentration was from 0.007 to 0.049 mg/L (Figure 6b).

Figure 6. Variation in pollution concentration and discharge during the period of study: (a) Biochemical Oxygen Demand (BOD); (b) Total Phosphorous (TP).

The comparison between pollution load and discharge on the corresponding days of 680 rainfalls is shown in Figure 7. The BOD load on the date of rainfall was in the range from 0.026 to 2374.5 ton/day (Figure 7a), while the TP load was from 0.59 to 4377.6 kg/day (Figure 7b). The relationships between discharge and BOD load, and between discharge and TP load were found to be very significant with the t -test result of $p < 0.001$.

Figure 7. Relationship between flow rate and pollution load: (a) BOD; (b) TP.

The relationship between discharge and BOD load was shown in Figure 7a, while the relationship between discharge and TP load was presented in Figure 7b.

Empirically, river discharge and pollution load have a relationship of $L = a \cdot Q^b$. Converting this into the form $L/Q = C = a \cdot Q^{(b-1)}$, where the concentration of the substance increases with the increase in discharge with an index b while the discharge coefficient is greater than 1, and decreases with b being less than 1 [42–45].

As shown in Figure 7a, the index b value of BOD was 0.93, which was less than 1, and the BOD concentration decreased with increased discharge, as in the aforementioned conditions. This is believed to be caused by the dilution effect due to the increase in discharge when it rains, because the BOD concentration is greatly influenced by the point source in general. As shown in Figure 7b, the index b value of TP was 1.12, which was greater than 1. This is believed to be caused by the effect of phosphorus, which is absorbed soil particles that flow to the waterbody by the increased discharge and rainfall.

3.2. Results of Input and Target Data Construction for CNN Model

During the data collection period, a total of 680 rainfalls occurred. Since the hydrological image, the input data of the CNN model, could be produced only with precipitation, a total of 680 hydrological images were produced in this study. In this regard, an example of hydrological image creation is demonstrated in Figure 8. Grid-type land use, soil, CN, and precipitation data with the same resolution (30×30 m) were used for the hydrological image creation, and the final hydrological images in the TIF format with the size of 956×633 grids were derived, as shown in Figure 8c. As mentioned previously in Section 2.3.1, if CN is allocated in grids, it resembles Figure 8a; with no rainfall conditions reflected, all input data will be the same, as shown in Figure 8a; but with rainfall conditions reflected, various hydrological images can be created, as shown in Figure 8c. When comparing Figure 8b,c, it becomes whiter as the precipitation amount increases and becomes darker as that amount decreases in Figure 8b. That is, as shown in Figure 8c, the brightness varies based on CN value and rainfall conditions according to Equations (1)–(3). Higher attribute values assigned to each grid are expressed in white and converge to zero as they get closer to black.

Figure 8. Example of hydrological image generation: (a) CN map; (b) precipitation map; (c) hydrological image; (c) calculation result using Equations (1)–(3); (d) actual hydrological image format as the model's input data.

Many previous studies have used 70 to 80% of the total input data to train a CNN or ANN model, and the rest as the test data to evaluate the model performance [46–49]; however, there are no

specific criteria for this. In this study, the data were classified as training, validation, and test datasets by shuffling (Table 4). A total of 501 out of 680 hydrological images were selected as the training and validation dataset (model training). Here, 400 were classified as the training dataset, while the remaining 101 were classified as the validation dataset. A total of 179 out of 680 hydrological images were set as the test dataset to predict BOD and TP loads for the whole study area (model evaluation).

Table 4. Classification of dataset.

Dataset		Number of Data			Remark
		Hydrological Image	BOD Load Target Data	TP Load Target Data	
Input dataset (model training)	Training data	400	400	400	Shuffled
	Validation data	101	101	101	
Test dataset (model evaluation)	Prediction data	179	179	179	
Total number of data		-	680	-	-

3.3. Results of Training for CNN Modeling

The results of the CNN model training are demonstrated in Figure 9. The learning of the Loss, Validation Loss, MAE and Validation MAE for BOD load are shown in Figure 9a–d; The learning of the Loss, Validation Loss, MAE and Validation MAE for TP load are presented in Figure 9e–h. When the CNN model is stably trained, the graph of the learning result exhibits an exponential function with a base of zero or less. It can be observed from Figure 9a–h that the learning progressed in a similar form to the exponential function graph, with all base values below zero.

Figure 9. Result of model training for the CNN model: (a–d) result of BOD load model; (e–h) result of TP load model.

As shown in Figure 9a,b,e,f, the Loss and MAE of BOD and TP load learning gradually decrease as the epoch progresses, indicating smooth progress of the model learning. The Validation Loss, together with Validation MAE of BOD and TP load learning, are shown in Figure 9c,d,g,h. The Validation Loss and Validation MAE indicate the generalization degree of the model, where, the generalization of the model is an indicator to evaluate the model performance for the new input data. Better generalization guarantees a higher model performance. As shown in Figure 9c,d,g,h, the optimization of Validation Loss and Validation MAE seems to be very unstable, but it tends to become more stable with repeated

learning, suggesting that the performance of the model gradually increases. Therefore, in this study, the model generalization for the testing stage seemed to have progressed based on the results of the Validation Loss and as the Validation MAE gradually decreases.

3.4. Prediction Result of CNN Model and Its Evaluation

The finally, for the prediction result and evaluation of the CNN model, the ped and measured values must be compared using the test dataset. This study used the previously constructed 179 test datasets to compare and analyze the actual measured values with the values predicted using the CNN model.

The scatter plot for the predicted and observed values of the model are presented in Figure 10. As shown in Figure 10a,b, the predicted and observed values have a positive relationship with each other, while the R^2 for the BOD load model was 0.88, which was higher than that for the TP load model, 0.75.

Figure 10. Comparison between observed and estimated values: (a) BOD; (b) TP.

The variation in the estimated and observed values for BOD and TP loads are given in Figure 11a,b. Upon looking at the variations in the estimated and observed values, the estimated value follows the trend of the observed value quite well. The BOD load model estimated somewhat lower loads than the observed BOD load among the high-level load; however, the TP load model estimated somewhat higher loads than the observed TP load among the high-level load.

Figure 11. Variation in the observed and estimated values: (a) BOD; (b) TP.

As suggested in this study, the Pearson correlation coefficient and NSE can be presented as the criteria for evaluating the model performance. Generally, when evaluating the model using the Pearson correlation coefficient value in the range from 0.1 to 0.3 indicates a weak performance, value from 0.4 to 0.6 indicates moderate performance, value from 0.7 to 0.9 indicates strong performance, and value of 1 indicates a perfect performance [50]. When using NSE as the criteria, if it is in the range of 0.6 to 0.8, it is evaluated as moderate to good, and it is evaluated as good [51] when above 0.8. Accordingly, if the result of this model is evaluated based on the Pearson correlation coefficient, the prediction performance for the BOD and TP loads can be regarded as strong. Based on NSE, the prediction performance for the BOD load can be evaluated as good, with the prediction performance for the TP load as moderate to good.

The Pearson correlation coefficient, NSE and MAPE between the predicted and measured values of BOD and TP loads are shown in Table 5. For the BOD load, the Pearson correlation coefficient, NSE and MAPE were 0.94, 0.83 and 53.346%, respectively, and for the TP load, they were 0.87, 0.79 and 75.621%, respectively.

Table 5. Model evaluation result.

Contents	<i>r</i>	NSE	MAPE (%)
BOD Load Predicted	0.94	0.83	53.346
TP Load Predicted	0.87	0.79	75.621

This study attempted to evaluate the model through comparison with other studies, but almost no similar studies were conducted previously; therefore, the results of previous studies on the prediction of pollutant load using ANN were referenced. Azadi et al. [52] used a model combining the Principal Component Analysis-M5P and ANN to simulate the COD load of landfill, and reported the superiority of the model with the Pearson correlation coefficient between simulated COD load and measured value of 0.98, and reported that the MAPEs for ANN and PCA-M5P models were 4% and 12%, respectively. Zhao et al. [53] predicted Ammonia nitrogen in Harbin using a model combining ANN and particle swarm optimization, and reported a good model performance with an NSE value higher than 0.6. Yu et al. [54] developed a combined kernel principal component analysis and extreme learning machine (KPCA-ELM) model to forecast the inlet chemical of sewage treatment, and reported a high accuracy for forecasting A and B, with an MAPE of 1.223% and 1.321%, respectively. Yan et al. [55] used a back-propagation neural network (BPNN), a genetic algorithm combined BPNN (GA-BPNN), a particle

swarm optimization combined with BPNN (PSO-BPNN), and a model combining PSO-BPNN and GA-BPNN(PSO-GA-BPNN) to predict the dissolved oxygen (DO). BPNN, GA-BPNN, PSO-BPNN, and PSO-GA-BPNN presented a good or excellent result in predicting DO, with an MAPE of 25.106%, 8.451%, 15.710%, and 6.722%, respectively. Shao et al. [56] forecasted daily water quality using the CS-BP model based on the backpropagation neural network optimized by the Cuckoo Search algorithm. The CS-BP model showed good forecast accuracy, with an MAPE of 0.001–2.689%.

Although the results of this study can be judged to have produced relatively good results based on the criteria of NSE and r , there is a limit to evaluating this study compared to the results of the MAPE in the preceding study. In this regard, the reason for the high MAPEs of this study could be found from two principal perspectives. The first reason is the absence of accurate pollution load data for hydrological images, which are an input dataset from the model. Typically, CNN models use the accurate train and target data. For example, CNN has used pictures of cats or dogs for model learning. However, as mentioned earlier, hydrological images do not have exact corresponding data. When model training, if the model uses data with very weak relationships between input and target data, this model only results in a trend with low accuracy, and the model cannot guarantee good results. As an alternative, the use of input and target data in relatively strong relationships can be suggested to improve this. The second reason is the structure of the CNN model or the hyperparameter. Prior studies reported good results by enhancing the structure or hyperparameters of CNN models to improve accuracy. We expect that this study can derive good results by modifying the CNN model structure, and adjusting hyperparameters such as learning rates, dropout, batch normalization, cross-validation.

As mentioned earlier, although the results of this study are difficult to judge as good, we found that the CNN using a hydrological image has the potential to utilize appropriate results in simulating pollution loads. If this study applies the improvements such as the above-mentioned method to enhance accuracy further, it could be applied to simulations of pollution loads other than BOD and TP loads.

This study demonstrates similar results to previous studies and we believe that the hydrological image has been well recognized by the CNN model, and led to proper results. Since the hydrological image's attribute value appears to play a sufficient role in predicting the pollutant load, it can be sufficiently applied to predict pollutants other than the BOD and TP loads simulated in this study.

4. Conclusions

This study aimed to develop a BOD and TP load prediction model using CNN, which can efficiently identify the spatial characteristics of input data among DNN algorithms, and to improve the change in the methodology of the pollution load prediction model by allowing the two-dimensional structure to be used as input data.

To this end, this study enabled the CNN model, through the convolution and pooling layer, to predict BOD and TP load using the hydrological image only and improved the data input method. The simulation results of BOD and TP load models revealed that all models' prediction performance turned out to be either moderate to good or good, and this could determine that it would be highly applicability in the field of prediction of pollution load. However, this study has the limitation of using only the CNN model's simplest structure to review the hydrological image's applicability as the input data.

Nevertheless, if this technology is gradually developed through this study's methodology or results, its application field could be extended, such as predicting loads in unmeasured watersheds, merging with remote sensing technology using satellite images and aerial photographs, or tracing the location of pollutant, etc. It is believed that the increased accuracy of predicting pollutant load could be derived through the improved CNN model or the complex structure CNN model.

Author Contributions: Conceptualization, C.M.S.; methodology, C.M.S.; software, C.M.S.; validation, J.S.K.; formal analysis, C.M.S.; investigation, C.M.S.; resources, C.M.S.; data curation, C.M.S., J.S.K.; writing—original, C.M.S.; draft preparation, C.M.S.; writing—review and editing, C.M.S., J.S.K.; visualization, C.M.S.; supervision, C.M.S., J.S.K.; project administration, C.M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Arnold, J.G.; Allen, P.M.; Bernhardt, G. A comprehensive surface-groundwater flow model. *J. Hydrol.* **1993**, *142*, 47–69. [CrossRef]
2. Aisha, M.S. Evaluation of SWAT Model Applicability for Water Impairment Identification and TMDL Analysis, The Degree of Doctor of Philosophy. Ph.D. Thesis, University of Maryland, University Park, MD, USA, 2007.
3. Young, R.A.; Onstad, C.A.; Bosch, D.D.; Anderson, W.P. AGNPS: A Nonpoint Source Pollution Model for Evaluating Agricultural Watersheds. *J. Soil Water Conserv.* **1989**, *44*, 168–173. Available online: https://www.researchgate.net/publication/277993805_AGNPS_A_Non-Point-Source_Pollution_Model_for_Evaluating_Agricultural_Watersheds (accessed on 13 March 2019).
4. Lewis, A.R. *Storm Water Management Model User's Manual*; Water Supply and Water Resources Division National Risk Management Research Laboratory: Cincinnati, OH, USA, 2004.
5. Patel, A.B.; Joshi, G.S. Modeling of Rainfall-Runoff Correlations Using Artificial Neural Network-A Case Study of Dharoi Watershed of a Sabarmati River Basin, India. *Civ. Eng. J.* **2017**, *3*, 78–87. [CrossRef]
6. Mercier, G.; Blais, J.F.; Chartier, M. Décontamination à l'échelle pilote de sols pollués en métaux toxiques par des procédés miniers et lixiviation chimique. *J. Environ. Eng. Sci.* **2007**, *6*, 53–64. [CrossRef]
7. Zhu, S.; Heddad, S. Prediction of dissolved oxygen in urban rivers at the Three Gorges Reservoir, China: Extreme learning machines (ELM) versus artificial neural network (ANN). *Water Qual. Res. J.* **2019**, *55*, 106–118. [CrossRef]
8. Rajae, T.; Khani, S.; Ravansalar, M. Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemom. Intell. Lab. Syst.* **2020**, *200*, 103978. [CrossRef]
9. Zhou, Y. Real-time probabilistic forecasting of river water quality under data missing situation: Deep learning plus post-processing techniques. *J. Hydrol.* **2020**, *589*, 125164. [CrossRef]
10. Pradhan, P.; Tingsanchali, T.; Shrestha, S. Evaluation of Soil and Water Assessment Tool and Artificial Neural Network models for hydrologic simulation in different climatic regions of Asia. *Sci. Total. Environ.* **2020**, *701*, 134308. [CrossRef]
11. Li, X.; Huang, M.; Wang, R. Numerical Simulation of Donghu Lake Hydrodynamics and Water Quality Based on Remote Sensing and MIKE 21. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 94. [CrossRef]
12. Wilby, R.; Abrahart, R.J.; Dawson, C. Detection of conceptual model rainfall—Runoff processes inside an artificial neural network. *Hydrol. Sci. J.* **2003**, *48*, 163–181. [CrossRef]
13. Jain, A.; Sudheer, K.P.; Srinivasulu, S. Identification of physical processes inherent in artificial neural network rainfall runoff models. *Hydrol. Process.* **2004**, *18*, 571–581. [CrossRef]
14. Sudheer, K.P.; Jain, A. Explaining the internal behaviour of artificial neural network river flow models. *Hydrol. Process.* **2004**, *18*, 833–844. [CrossRef]
15. Van, S.P.; Le, H.M.; Thanh, D.V.; Dang, T.D.; Loc, H.H.; Anh, D.T. Deep learning convolutional neural network in rainfall—runoff modelling. *J. Hydroinform.* **2020**, *22*, 541–561. [CrossRef]
16. Li, X.; Du, Z.; Song, G. A Method of Rainfall Runoff Forecasting Based on Deep Convolution Neural Networks. In Proceedings of the 2018 Sixth International Conference on Advanced Cloud and Big Data (CBD), Lanzhou, China, 12–15 August 2018; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2018; pp. 304–310.
17. KMA: Korea Meteorological Administration. Available online: <https://www.kma.go.kr> (accessed on 3 January 2020).
18. WAMIS: Water Management Information System, National Institute of Environmental Research. Available online: <https://www.water.nier.go.kr> (accessed on 1 March 2019).

19. EGIS: Environmental Geographic Information Service. Available online: <https://www.egis.me.go.kr> (accessed on 9 January 2019).
20. NIER: National Institute of Environmental Research. Available online: <https://www.nier.%20go.kr> (accessed on 9 January 2019).
21. Song, C.M. Hydrological Image Building Using Curve Number and Prediction and Evaluation of Runoff through Convolution Neural Network. *Water* **2020**, *12*, 2292. [CrossRef]
22. Li, C.; Liu, M.; Hu, Y.; Shi, T.; Zong, M.; Walter, M.T. Assessing the Impact of Urbanization on Direct Runoff Using Improved Composite CN Method in a Large Urban Area. *Int. J. Environ. Res. Public Heal.* **2018**, *15*, 775. [CrossRef]
23. Wang, H.; Chen, Y. Identifying Key Hydrological Processes in Highly Urbanized Watersheds for Flood Forecasting with a Distributed Hydrological Model. *Water* **2019**, *11*, 1641. [CrossRef]
24. Ministry of Land, Infrastructure and Transport, South Korea. *Design Flood Estimation Techniques*; Ministry of Land Transport and Maritime Affairs: Seoul, South Korea, 2012. (In Korean)
25. Eurostat. Available online: http://ec.europa.eu/eurostat/statistics-explained/index.php/People_at_risk_of_poverty_or_social_exclusion (accessed on 6 June 2020).
26. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
27. Medina, E.; Petraglia, M.R.; Gomes, J.G.R.C.; Petraglia, A. Comparison of CNN and MLP classifiers for algae detection in underwater pipelines. In Proceedings of the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada, 28 November–1 December 2017; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2017; pp. 1–6.
28. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. In *Advances in Cryptology—EUROCRYPT 2018*; Springer: Cham, Switzerland, 2014; Volume 8689, pp. 818–833. [CrossRef]
29. Keras. Available online: <https://keras.io> (accessed on 12 December 2019).
30. Tensorflow. Available online: <https://www.tensorflow.org> (accessed on 12 December 2019).
31. Ide, H.; Kurita, T. Improvement of learning for CNN with ReLU activation by sparse regularization. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 2684–2691. [CrossRef]
32. Chen, Z.; Ho, P.-H. Global-connected network with generalized ReLU activation. *Pattern Recognit.* **2019**, *96*, 106961. [CrossRef]
33. Bottou, L. Large-Scale Machine Learning with Stochastic Gradient Descent. In Proceedings of the COMPSTAT'2010, Paris, France, 22–27 August 2010; Springer Science and Business Media LLC: Heidelberg, Germany, 2010; pp. 177–186.
34. Qian, N. On the momentum term in gradient descent learning algorithms. *Neural Netw.* **1999**, *12*, 145–151. [CrossRef]
35. Nesterov, Y. A method for unconstrained convex minimization problem with the rate of convergence. *Doklady Ussr* **1983**, *269*, 543–547. Available online: <https://www.semanticscholar.org/paper/A-method-for-unconstrained-convex-minimization-with-Nesterov/ed910d96802212c9e45d956adaa27d915f5d7469> (accessed on 18 April 2020).
36. Duchi, J.; Hazan, E.; Singer, Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *JMLR* **2011**, *12*, 2121–2159. Available online: <https://www.jmlr.org/papers/v12> (accessed on 11 May 2020).
37. Dozat, T. *Incorporating Nesterov Momentum into Adam*; ICLR Workshop: San Juan, Puerto Rico, 2016; Available online: <https://openreview.net/pdf/OM0jvwB8jlp57ZJjtNEZ.pdf> (accessed on 21 April 2020).
38. Zeiler, M.D. ADADELTA: An Adaptive Learning Rate Method. Available online: <https://arxiv.org/pdf/1212.5701.pdf> (accessed on 21 May 2020).
39. Hinton, G.; Tieleman, T. RMSprop Gradient Optimization; Lecture 6e of his Coursera Class. 2014. Available online: https://www.cs.toronto.edu/~tieleman/protect/T1\textbraceleft\protect/T1\textbracerighttjmen/csc321/slides/lecture_slides lec6.pdf (accessed on 10 February 2019).
40. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
41. Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 1st ed.; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2017; ISBN 97981491962299.
42. Gunnerson, C.G. *Stream Flow and Quality in the Columbia River Basin*; ASCE: Reston, VA, USA, 1967; p. 93.

43. Langmuir, D. *Aqueous Environmental Geochemistry*; Prentice-Hall Inc.: Upper Saddle River, NJ, USA, 1997; p. 600.
44. Appelo, C.; Postma, D. *Geochemistry, Groundwater and Pollution*; Informa UK Limited: Colchester, UK, 2004; p. 536.
45. Igarashi, T.; Hataya, R.; Oyama, T. Estimation of pyrite oxidation rate by sulfate ion discharged from a catchment. *J. Geochem. Explor.* **2003**, *77*, 151–165. [[CrossRef](#)]
46. Lee, K.; Choi, C.; Shin, D.H.; Kim, H.S. Prediction of Heavy Rain Damage Using Deep Learning. *Water* **2020**, *12*, 1942. [[CrossRef](#)]
47. Alsumaiei, A.A. Utility of Artificial Neural Networks in Modeling Pan Evaporation in Hyper-Arid Climates. *Water* **2020**, *12*, 1508. [[CrossRef](#)]
48. Lee, J.; Kim, C.-G.; Lee, J.E.; Kim, N.W.; Kim, H. Medium-Term Rainfall Forecasts Using Artificial Neural Networks with Monte-Carlo Cross-Validation and Aggregation for the Han River Basin, Korea. *Water* **2020**, *12*, 1743. [[CrossRef](#)]
49. Mulualem, G.M.; Liou, Y.-A. Application of Artificial Neural Networks in Forecasting a Standardized Precipitation Evapotranspiration Index for the Upper Blue Nile Basin. *Water* **2020**, *12*, 643. [[CrossRef](#)]
50. Dancy, C.; Reidy, J. *Statistics without Maths for Psychology*, 5th ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2011.
51. Lipiwattanakarn, S.; Saengsawang, S. Performance comparison of a conceptual hydrological model and a back-propagation neural network model in rainfall-runoff modeling. *Eng. J. Res. Dev.* **2005**, *16*, 35–42. Available online: <https://ph02.tci-thaijo.org/index.php/eit-researchjournal/article/view/147563/108653> (accessed on 13 July 2020).
52. Azadi, S.; Amiri, H.; Rakhshandehroo, G.R. Evaluating the ability of artificial neural network and PCA-M5P models in predicting leachate COD load in landfills. *Waste Manag.* **2016**, *55*, 220–230. [[CrossRef](#)]
53. Zhao, Y.; Guo, L.; Wang, Y.; Wang, P. Regional optimal allocation for reducing waste loads via artificial neural network and particle swarm optimization: A case study of ammonia nitrogen in Harbin, northeast China. *Water Sci. Technol.* **2014**, *70*, 1211–1219. [[CrossRef](#)]
54. Yu, T.; Yang, S.; Bai, Y.; Gao, X.; Li, C. Inlet Water Quality Forecasting of Wastewater Treatment Based on Kernel Principal Component Analysis and an Extreme Learning Machine. *Water* **2018**, *10*, 873. [[CrossRef](#)]
55. Yan, J.; Xu, Z.; Yu, Y.; Xu, H.; Gao, K. Application of a Hybrid Optimized BP Network Model to Estimate Water Quality Parameters of Beihai Lake in Beijing. *Appl. Sci.* **2019**, *9*, 1863. [[CrossRef](#)]
56. Shao, D.; Nong, X.; Tan, X.; Chen, S.; Xu, B.; Hu, N. Daily Water Quality Forecast of the South-To-North Water Diversion Project of China Based on the Cuckoo Search-Back Propagation Neural Network. *Water* **2018**, *10*, 1471. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).